भारतीय आधुनिक शिक्षा राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद् की एक त्रैमासिक पत्रिका है। इस पत्रिका का मुख्य उद्देश्य है शिक्षकों, शिक्षक-प्रशिक्षकों, शैक्षिक प्रशासकों तथा शोधकर्ताओं को एक मंच प्रदान करना, शिक्षा के विभिन्न आयामों जैसे–शिक्षादर्शन, शिक्षा मनोविज्ञान, शिक्षा की समकालीन समस्याएं, पाठ्यक्रम एवं प्रविधि संबंधी नवीन विकास, अंतरराष्ट्रीय स्तर पर शिक्षा का स्वरूप, विभिन्न राज्यों में शिक्षा की स्थिति आदि पर मौलिक तथा आलोचनात्मक चिंतन को प्रोत्साहित करना और शिक्षा के क्षेत्र में सुधार और विकास को बढ़ावा देना।

लेखकों द्वारा व्यक्त किए गए विचार उनके अपने हैं तथा ये किसी भी प्रकार परिषद् की नीतियों को प्रस्तुत नहीं करते।

प्रकाशन सहयोग

पूरन चंद *अध्यक्ष,* प्रकाशन प्रभाग

मीरा कांत *संपादक*

कल्याण बैनर्जी *उत्पादन अधिकारी*

अरुण चितकारा *सहायक उत्पादन अधिकारी*

राजेश पिप्पल *उत्पादन सहायक*

मूल्य एक प्रति : 8.50 रुपये; वार्षिक : 34.00 रुपये

भारतीय आधुनिक शिक्षा

वर्ष : 17 | अंक : 2 | अक्तूबर 1999

विषय सूची

भारत की गौरवशाली गुरु-शिष्य परंपरा

हीरालाल बाछोतिया

के 40 एफ साकेत
नई दिल्ली-110017

शिक्षा का प्रमुख उद्देश्य परंपरा की धरोहर को एक पीढ़ी से दूसरी पीढ़ी तक पहुंचाना है। इस प्रक्रिया में परंपरा का रचनात्मक मूल्यांकन भी शामिल है। किंतु आज शिक्षा संस्थाओं में जो कुछ हो रहा है वह चिंतनीय है। आधुनिकीकरण की भ्रमपूर्ण संकल्पना के कारण नई पीढ़ी धुरीहीन हो रही है। गुरु-शिष्य संबंधों में तनाव देखा जा रहा है। किंतु यह प्रतीति भी जनमानस में है कि गुरु का स्थान निर्माता का है। दोनों के संबंधों के बारे में शिक्षाविदों ने एक नीति संहिता तैयार की है जो गुरु-शिष्य संबंध को आधुनिक परिप्रेक्ष्य में व्यावहारिक बनाने की प्रक्रिया ही है।

गुरु के प्रति आदर भारत की एक जीवंत परंपरा है। गुरु को ईश्वर से बढ़कर बताया गया है। भारतीय वाङ्मय में ज्ञान को तीसरा नेत्र माना गया है। वैदिक दार्शनिकों ने ज्ञान को सत्य, अनंत और ब्रह्म माना है। यह ज्ञान बिना गुरु के संभव नहीं है इसीलिए कबीर कहते हैं –

> गुरु गोविंद दोऊ खड़े, काके लागूं पांय।
> बलिहारी गुरु आपने जिन गोविंद दियो बताय।

गुरु-शिष्य संबंध पिता-पुत्र जैसा माना गया है। यह अत्यंत पवित्र और निकट का संबंध है। निश्चय ही इसमें एक ओर आदर, श्रद्धा और दूसरी ओर स्नेह का भाव सन्निहित है। ऐसे वातावरण में विद्याध्ययन एक यज्ञ के समान हो जाता था। विद्याध्ययन या ज्ञानार्जन को यज्ञ ही माना गया है। शिष्य में शुद्धता, ज्ञान की पिपासा, लगन, परिश्रम तथा साधना का होना आवश्यक है।

वैदिक साहित्य में छात्र के लिए 'आचार्य कुलवासी' और 'अंतेवासी' शब्दों का प्रयोग मिलता है। यह इस बात का संकेत करता है कि शिक्षार्थी को गुरु के आश्रम में रहकर ही शिक्षा ग्रहण करनी है। गुरु के आश्रम में शिष्य को वेद एवं वेदेतर विषयों की मौखिक शिक्षा के साथ-साथ सामाजिक व वैयक्तिक अनुशासन, सत्य-निष्ठा और अपने कर्तव्यों के प्रति सतत जागरूक रहने की शिक्षा दी जाती थी। गुरुकुल में अपनी शिक्षा समाप्त करने के बाद स्नातक न केवल अपनी और अपने प्रतिपाल्य जनों की सारी ज़िम्मेदारियां संभाल लेता था वरन समाज के प्रति अपने उत्तरदायित्व का बोध भी बड़ी कुशलता के साथ निभाता था।

वैदिक सूक्त का आशय है कि यदि कोई योद्धा युद्ध में शौर्य प्रदर्शित करता है, यदि कोई राजा प्रशासन में सफल होता है, यदि कोई व्यापारी अच्छा लाभ कमाता है तो यह उसके छात्र जीवन की शिक्षा का ही सुपरिणाम होता है। आशय यह कि शिक्षा पूरे जीवन की नियामक है। इस शिक्षा के आधार-स्तंभ गुरु ही होते थे। भारतीय परंपरा में गुरु को सर्वोच्च स्थान प्राप्त है। यह शायद इसलिए भी कि गुरु को परंपरा का हस्तांतरणकर्ता निरूपित किया गया है। उसके कंधों पर परंपरा निर्वाह का संपूर्ण भार डाल दिया गया है। यह दायित्व समय, परिस्थिति और देशकाल के संदर्भ में उचित ही था। आज भी उससे ऐसे दायित्व की आशा की जाती है। स्वामी विवेकानंद के अनुसार 'सही गुरु तो वह है जो अपनी आत्मा को शिष्य की आत्मा में प्रविष्ट करा दे तथा शिष्य के मन द्वारा देख और समझ सके।' स्वामी विवेकानंद के कथन में गुरु का आध्यात्मिक निरूपण अधिक प्रतीत हो सकता है किंतु शिष्य के मन की बात जान लेना तो गुरु के लिए आवश्यक ही है जिसके अभाव में अध्यापन एकपक्षीय हो सकता है।

शिष्य के लिए गुरु एक आदर्श होता है। उसकी विचारधारा, रहन-सहन और सबसे बढ़कर आचरण का प्रभाव शिष्य पर और प्रकारांतर से समाज पर पड़ता है। गुरुदेव रवींद्रनाथ टैगोर के अनुसार अध्यापक को संयमी, ज्ञानी और आदर्श की प्रतिमूर्ति होना चाहिए। अध्यापक के ज्ञान की अपेक्षा उसके आचरणों और मनोभावों का प्रभाव शिष्य पर पड़ता है। स्पष्ट है कि सच्चरित्र अध्यापक देश की अमूल्य निधि है। आज नैतिक और जीवन मूल्यों की शिक्षा की चर्चा की जाती है। इस दृष्टि से भी गुरु या शिक्षक अपने आचरण द्वारा ही नैतिक और जीवन मूल्यों की शिक्षा को मूर्त रूप दे सकता है।

महर्षि अरविंद गुरु को शिष्य के पथद्रष्टा एवं सहायक के रूप में देखते हैं। उनका मत है कि मानव का ज्ञान उसके अंदर छिपा हुआ है और शिक्षक या गुरु उसके आवरण को हटा देता है। आशय यह कि शिक्षा का केंद्र बालक या शिष्य है। शिक्षक न तो बालकों को ज्ञान प्रदान करता है और न अंतर्निहित ज्ञान को विकसित करता है अपितु छात्रों की इस प्रक्रिया में सहायता करता है कि वे अपने अंदर निहित ज्ञान का विकास करें। बालक आधारित शिक्षा में शिक्षक को गौण नहीं बनाया गया है वरन उसे महत्तर भूमिका दी गई है। बालक की क्षमताओं को पहचानना निश्चय ही एक महत्वपूर्ण कार्य है। इस दृष्टि से शिक्षक में एक अंतर्दृष्टि की आशा की गई है।

शिक्षक के थोड़े से प्रोत्साहन से भी बालक में अपूर्व आत्मविश्वास का विकास हो सकता है। आत्मविश्वास के द्वारा बालक अपनी लगन, परिश्रम और साधना से वह सब कुछ प्राप्त कर दिखाता है। उसे देखकर अपार प्रसन्नता या तो माता-पिता को होती है या शिक्षक को, इस अंतर के साथ कि शिक्षक की यह प्रसन्नता निःस्वार्थ भाव से प्रेरित होती है। भारतीय दृष्टि में शिक्षा एक दायित्व है। व्यक्ति पर शिक्षा एक ऋण है, जो सीखा है वह आने वाली पीढ़ी को सिखाना है। यह गुरु ऋण है जिससे उऋण होना है। इसके लिए जो कुछ प्राप्त किया है उसमें कुछ जोड़ना भी होगा। यह है

अपना अनुभव। यह कोई उपकार नहीं वरन दायित्व भाव से शिक्षक इसे करता रहा है। पूर्वजों से प्राप्त धरोहर को आगे वाली पीढ़ी को सौंपता रहा है।

शिक्षा सामाजिक प्रक्रिया है जो अपने परिवेश को संजोए हुए है। अनुभव के साथ ज्ञान की थाती में वृद्धि होती रहती है। जब कालांतर में पाश्चात्य शिक्षा से भारतीय शिक्षा चिंतन का आमना-सामना हुआ तो अनेक प्रश्न उपस्थित हो गए। पाश्चात्य दृष्टिकोण का उद्देश्य बालक को केंद्र में रखना था। पाश्चात्य आदर्शवादी शिक्षा-दर्शन में भी बालक केंद्रित शिक्षा पर बल दिया गया है। जॉन डी.वी. का कहना है कि शिक्षा की प्रक्रिया समायोजन की अनवरत प्रक्रिया है जिसका लक्ष्य हर अवस्था में बुद्धिगत सामर्थ्य के रूप में होता है। कहना न होगा कि भारतीय परंपरा से पाश्चात्य चिंतन का टकराव अवश्यंभावी था किंतु खास बात यह रही है कि भारतीय शिक्षा दर्शन ने सबके साथ समायोजन किया और अपने आपको परंपरा से कटने भी नहीं दिया।

स्मरणीय है कि पराधीनता ने हमारे शिक्षा के संस्कारों पर चोट करने की कोशिश भी की। भारतीयों में हीनताबोध उत्पन्न करना चाहा। किंतु स्वामी दयानंद, विवेकानंद, अरविंद, महात्मा गांधी जैसे राष्ट्र निर्माताओं ने हमारे बोध को ललकारा। शिक्षक ने फिर अपने दायित्व का निर्वाह किया। अपने शिष्यों में राष्ट्रीयता की भावना का अंकुरण किया। महात्मा गांधी ने बुनियादी तालीम के माध्यम से तो गुरुदेव रवींद्रनाथ ठाकुर ने प्रकृति की गोद में शिक्षा का संदेश देकर गुरु-शिष्य संबंध को भारतीय परंपरा से जोड़ा। इस प्रकार शिक्षा द्वारा समाज को संस्कारित करने की धारणा मूर्त रूप में प्रवाहित हो उठी जिसमें अवगाहन करने का अवसर हर एक के लिए उपलब्ध हुआ।

आज ऐसी शिक्षा की ज़रूरत है जो समाज और देश के लिए समर्पित विद्यार्थियों का निर्माण कर सके। यह गुरुतर कार्य गुरु के सिवाय और कौन करेगा। आज चारों ओर से चुनौतियां उपस्थित हो रही हैं। चुनौतियों की इन दुर्दांत लहरों के बीच से नई पीढ़ी की नाव को दिशा देने का काम शिक्षक ही कर सकता है। दूरस्थ शिक्षा के शिक्षार्थी भी किसी-न-किसी स्तर पर शिक्षक के सान्निध्य में रहते हैं। परिस्थितिवश पढ़ाई छोड़ देने के बाद दूरस्थ शिक्षा के माध्यम से अपनी आकांक्षाओं को पंख लगाते हैं। शिक्षक का कोई न कोई रूप उनके सामने भी होता ही है।

पीढ़ियों से संचित ज्ञान को आत्मसात कर उसे सुबोध बना कर छात्र या शिष्य को और इस प्रकार समाज को सौंप देना जब वैयक्तिक स्तर पर संभव नहीं रह गया तो शिक्षा का संस्थानिक रूप सामने आया। शिक्षण संस्थाएं अस्तित्व में आईं। वास्तव में तभी गुरु 'शिक्षक' बना। राष्ट्र के मानस के अनुरूप ज्ञान की इस थाती को एक रूप मिला। समाज की आशाएं-आकांक्षाएं भी इसमें जुड़ती गईं।

भारतीय शिक्षा शास्त्री शिक्षा को व्यक्ति के शारीरिक, मानसिक, बौद्धिक एवं आत्मिक विकास की प्रक्रिया मानते हैं। शिक्षा ऐसा साधन है जो व्यक्ति तथा समाज को प्रगति एवं विकास की गति प्रदान करता है। भारतीय आचार्यों ने शिक्षक को शिक्षा का एक ध्रुव माना है तो शिष्य को दूसरा

ध्रुव। गुरु शिक्षा का पूर्व रूप है तो छात्र उसका उत्तर रूप। पाश्चात्य शिक्षा शास्त्रियों एडम्स और रॉस ने भी शिक्षा को ही ध्रुवीय प्रक्रिया माना है। अतः गुरु-शिष्य अथवा शिक्षक-शिक्षार्थी संबंधों की भारतीय परंपरा को आज के परिप्रेक्ष्य में भी जीवंत बनाए रखने की आवश्यकता पहले से भी अधिक है।

कहना न होगा कि शिक्षा का प्रमुख उद्देश्य परंपरा की धरोहर को एक पीढ़ी से दूसरी पीढ़ी तक पहुंचाना है। इस प्रक्रिया में परंपरा का रचनात्मक मूल्यांकन भी शामिल है। किंतु हमारी शिक्षा संस्थाओं में भारतीयता की तलाश की जगह कुछ और हो रहा है। इन संस्थाओं में खिचड़ी संस्कृति बहुत अंशों में पश्चिम की नकल के रूप में प्रोत्साहित की जा रही है। यह शिक्षा नई पीढ़ी और परंपरा में अलगाव पैदा कर रही है। जनसाधारण से उसकी दूरी बढ़ती जा रही है। आधुनिकीकरण की भ्रमपूर्ण संकल्पना के कारण नई पीढ़ी धुरीहीन हो रही है। वह न अपनी परंपरा से पोषण पा रही है, न उसमें पश्चिम की सांस्कृतिक विशेषताएं ही आ रही हैं। शिक्षक-शिष्य के संबंधों में तनाव आ रहा है। शिक्षा प्रक्रिया पर परीक्षाएं हावी हो गई हैं। ये परीक्षाएं विद्यार्थी की किन क्षमताओं और कौशलों का मूल्यांकन करती हैं, कहना मुश्किल है। हां यह अवश्य हुआ है कि इनके कारण शिक्षा की महती भूमिका अवश्य प्रभावित हो रही है। विशेष रूप से शिक्षक या भारतीय परंपरा में जिसे गुरु के उच्चतम स्थान पर हमने स्वीकार किया था उसकी भूमिका प्रश्नों के घेरे में है। तथापि भारतीय जनमानस में गुरु 'मासशाब' या शिक्षक भले बन गया हो पब्लिक स्कूल का 'टीचर' नहीं बना है। जनमानस में शिक्षक के प्रति आज भी यह भाव तो है ही कि वह उसके बच्चे का भविष्य संवार सकता है। कबीर के शब्दों में कहें –

> गुरु कुम्हार शिस कुंभ है, कढ़ि-कढि काढ़े खोट।
> अंतर हाथ सराहि दे बाहर मारे चोट।।

अर्थात यदि शिक्षक कुम्हार की तरह शिष्य रूपी घड़े की ऊपर से पिटाई करता है तो अंदर से सहारा भी देता है। भारतीय जनमानस अभी इतना अनुदार नहीं हुआ है कि रूप देने वाले और भविष्य संवारने वाले को निरा वेतनभोगी माने। जनमानस आज भी गुरु नहीं तो शिक्षक को सिर आंखों पर बैठाए हुए है। यही भारतीय परंपरा है, भारत की गौरवशाली गुरु-शिष्य परंपरा। अभी हाल ही में प्राथमिक और माध्यमिक विद्यालयों के अध्यापक संघों के अखिल भारतीय महासंघ और शिक्षाविदों के दल ने अध्यापक-छात्र संबंध के बारे में एक नीति संहिता तैयार की। इसके अंतर्गत अध्यापकों का छात्रों से कैसा संबंध हो, इसे लेकर कुछ बिंदु तय किए गए। इसके अनुसार अध्यापकों को चाहिए कि वे –

- ❑ विद्यालय में अपने कर्तव्य को निभाने के लिए सदैव समय के पाबंद रहें।
- ❑ पढ़ाए जाने वाले पाठों को पूरी तरह से तैयारी करने के बाद ही पढ़ाएं।
- ❑ सभी बच्चों को एक जैसा प्यार तथा स्नेह दें और बिना किसी जाति, वर्ग, रंग, लिंग, प्रतिष्ठा, धर्म, भाषा और जन्म स्थान के भेदभाव के, सबके लिए एक समान रहें।

- विद्यार्थियों के शारीरिक, सामाजिक, बौद्धिक, भावनात्मक, नैतिक और आध्यात्मिक विकास में मार्गदर्शक बनें।
- विद्यार्थियों की व्यक्तिगत आवश्यकताओं और विद्यार्थियों की सामाजिक-सांस्कृतिक पृष्ठभूमि की ओर ध्यान दें तथा अपनी अध्यापन शैली को उसके अनुरूप ढालें।
- उपचारात्मक शिक्षण के लिए अनुमोदित किसी योजना के अतिरिक्त अपने विद्यार्थियों को अतिरिक्त समय में पढ़ाने अथवा ट्यूशन देने के लिए कोई राशि स्वीकार करने से बचें।
- उस सूचना के अतिरिक्त जो वैधानिक रूप से आवश्यक हो, उन्हें विद्यार्थियों के संबंध में किसी गोपनीय सूचना का रहस्योद्घाटन नहीं करना चाहिए।
- विद्यार्थियों को अन्य विद्यार्थियों, अध्यापकों अथवा प्रशासन के विरुद्ध उकसाने का प्रयास नहीं करना चाहिए।
- ऐसा पहनावा, बोलचाल और व्यवहार करें जो विद्यार्थियों के लिए आदर्श बन सके।
- विद्यालय में अनुशासन का पालन करते समय मूलभूत मानवीय गरिमा को सम्मान दें।

कहना न होगा कि यह संहिता गुरु-शिष्य संबंध को आधुनिक परिप्रेक्ष्य में व्यावहारिक बनाने की एक प्रक्रिया ही है। समग्र समाज और राष्ट्र की अपेक्षानुरूप गुरु-शिष्य संबंध का उच्च आदर्शों से संचालित होना अवश्यंभावी है। गुरु को 'गोविंद' नहीं तो राष्ट्र निर्माता तो स्वीकार करना ही होगा, तभी शिक्षा के ये मंदिर कर्म, श्रम और ज्ञान की किरणें बिखेरते रह सकेंगे। ❑

संस्कृति बनाम बहुसंस्कृति : शिक्षा के संदर्भ में

विजय कुमार सुनवानी

रीडर, अंग्रेज़ी
क्षेत्रीय शिक्षा संस्थान, भुवनेश्वर

भारत की सभ्यता बेबीलोन, मेसोपोटामिया जैसी पुरानी सभ्यताओं के समान श्रेष्ठ है। ऐसी सभ्यता, ऐसी संस्कृति पर गर्व होना स्वाभाविक है। यह दुखद है कि प्रायः लोग संस्कृति एवं सभ्यता जैसे गंभीर विषयों पर हल्के-फुल्के वक्तव्य देने में नहीं हिचकते और इस प्रकार संस्कृति के उच्चमान को कम करते हैं। हम सब को संस्कृति, सभ्यता के बारे में गहरा चिंतन करना चाहिए। प्रस्तुत लेख में संस्कृति और उससे जुड़े कुछ विषयों पर चर्चा की गई है। बहुभाषी, बहुसंस्कृतीय, बहुधार्मिकता प्रधान देश में औपचारिक और अनौपचारिक शिक्षा से हम किस प्रकार लाभान्वित हो सकते हैं, इस पर भी कुछ सुझाव प्रस्तुत किए गए हैं।

आजकल संस्कृति के बारे में यदा-कदा चर्चाएं होती ही रहती हैं। भारतीय संस्कृति को अपने ही देश में बहुत सराहा जा रहा है। उस पर किसी प्रकार का कुप्रभाव न पड़े, ऐसा सोचा जा रहा है। उसे दूसरे विदेशी चैनलों के बुरे प्रभावों से दूर रहना चाहिए ताकि हमारे बच्चे इन कुप्रभावों से बचे रहें। कोका कोला, पेप्सी और बार्बी की संस्कृति विदेशी है, खराब है। हमारी संस्कृति पवित्र, शुद्ध एवं स्वच्छ है। पर क्या भारतीय संस्कृति इतनी कमज़ोर है कि इन छोटी-छोटी बातों से मैली हो जाएगी? क्या वह इतनी जीर्ण हो गई है कि एम.टी.वी., जीन्स और चैनेल वी के शोर-गुल उसे प्रभावित कर सकें? जी नहीं। हम सब आजकल संस्कृति के बारे में कुछ ज़्यादा ही शोर-गुल कर रहे हैं। पर संस्कृति क्या है, हमें यह पहले अच्छी तरह समझ लेना चाहिए।

ऐसा लगता है कि हम संस्कृति को बहुत ही संकीर्ण मापदंड से परिभाषित कर रहे हैं। इसलिए हमें उसे समग्रता में परिभाषित करना होगा एवं समूचे समाज के परिवेश में देखना होगा कि भारतीय संस्कृति क्या है। यह सोचना आवश्यक है कि भारत एक विशाल देश है जहां अलग-अलग भाषाएं बोली जाती हैं, विभिन्न प्रांतों में किस्म-किस्म के लोग निवास करते हैं, जीवन-यापन करते हैं। साथ ही ये लोग अपनी-अपनी संस्कृति का अनुपालन करते हैं। ऐसा करते हुए भी हम सब की संस्कृति भारतीय है और वह क्षेत्रीय, प्रांतीय संस्कृतियों का समुच्चय है।

अपनी संस्कृति पर गर्व होना स्वाभाविक है परंतु इसका यह मतलब नहीं कि दूसरे की संस्कृति

खराब है, हमारी संस्कृति से हेय है। यह बात दूसरे देश की संस्कृति के लिए भी उतनी ही सही है जितनी कि भारत के विभिन्न प्रांतों के लिए। इसलिए हर देश की, हर प्रांत की, सबकी संस्कृति का आदर होना चाहिए। वस्तुतः दूसरे की संस्कृति और हमारी संस्कृति में किसी भी प्रकार की तुलनात्मक चर्चा होनी ही नहीं चाहिए। हर एक की संस्कृति अपने में श्रेष्ठ है और इसलिए हर संस्कृति का समान भाव से सम्मान होना चाहिए। साथ ही हमारे देश में बहुत-सी अनुसूचित जनजातियां भी हैं, इन सभी का अपने-अपने समुदायों में गौरवपूर्ण स्थान है, महत्व है। हमें यह भी नहीं भूलना चाहिए कि ये सब जनजातियां अपनी संस्कृति, अपने मूल्य, अपने मापदंड के बारे में गर्व का अनुभव करती हैं।

संस्कृति जीवन जीने की शैली है जो कि वर्ग विशेष द्वारा अपनाई गई होती है। यह विचारधारा, रहन-सहन, मूल्य, भाषा, दृष्टिकोण, राजनैतिक संगठन, अर्थनीति आदि का नाम है जो एक पीढ़ी से दूसरी पीढ़ी को सिखाई जाती है। संस्कृति हमें जीवन जीने एवं उसे समझने के मूलभूत सिद्धांत सिखाती है।

संस्कृति भाषा का एक अभिन्न अंग है, इन्हें एक-दूसरे से अलग नहीं किया जा सकता है। एक प्रकार से भाषा-शिक्षण, संस्कृति-शिक्षण ही है, परंतु भाषा-शिक्षण के माध्यम से सांस्कृतिक साम्राज्यवाद लाने का प्रयास उचित नहीं होता। अंग्रेज़ी के आधिपत्य के बारे में भी बहुत कुछ कहा जा चुका है - ये गुलामी की भाषा है, गुलामी की प्रतीक है जिसने हमारे मस्तिष्क तथा सृजन शक्ति को वशीभूत कर लिया है। हमें नहीं भूलना चाहिए कि जो भी देश किसी और देश पर आक्रमण करता है एवं जीत लेता है, उसे ऐसी सुविधा मिल जाती है कि वह अपनी भाषा को दूसरे देश में प्रतिष्ठित करे। ऐसी स्थिति में यदि हम किसी दूसरे देश के अधीन होते तो शायद हम उस देश की भाषा/भाषाओं को भी सीख गए होते। फिर अंग्रेज़ी का चयन इसलिए भी है कि यही एक भाषा है जो कि तटस्थ है - हमारी 18 भाषाओं के बावजूद जो कि संविधान द्वारा मान्यताप्राप्त हैं। आज के संदर्भ में ऐसी धारणा बन ही गई है, कि संस्कृत से आप विद्वान माने जाते हैं, हिंदी से आपकी राष्ट्रीयता बुलंद दिखती है, पर अंग्रेज़ी ने शायद आजकल खोखली बातचीत की भाषा का संशयात्मक दर्जा प्राप्त कर लिया है। भाषा और धर्म, संस्कृति से अवश्य जुड़े हुए हैं, पर इससे यह ग़लत निष्कर्ष निकाल लिया गया कि अंग्रेज़ी भाषा ईसाई समुदाय से जुड़ी हुई है। वास्तव में, अंग्रेज़ी एक तटस्थ भाषा थी, पर अब शायद यह भारत में इस रूप में नहीं मानी जाती क्योंकि संभवतः इस समय यहां भाषाएं अंतःकरण तक पहुंच कर भावावेश जगाती हैं।

हम लोग स्वयं ही अंग्रेज़ी भाषा को यहां रखने में हर स्तर पर लगे हुए हैं। यहां तक कि छोटे-छोटे ढाबों में यूरोप के एवं चायनीज़ खाने का प्रचार करना व मिल भी जाना आम बात होने लगी है। हिंदी फिल्मों के नामों में भी अंग्रेज़ी का उपयोग रुतबे की बात मानी जाती है। प्रायः सभी जगह अंग्रेज़ी वर्तनी में भूल मिलेगी। न ठीक से लिख सके, न बोल सके, न समझ सके, फिर भी न छूटे अंग्रेज़ी, वाह रे वाह!

भारत में सिर्फ 18 भाषाओं को लेकर नहीं चलना चाहिए। वास्तव में भारत में 1682 मातृ भाषाएं हैं, जो एक-दूसरे से एवं 18 मान्यता प्राप्त भाषाओं से एकदम भिन्न हैं। इसीलिए मातृभाषा और विद्यालय भाषा में फर्क होना संभव है। ऐसा होने पर यदि छात्र को विद्यालय की भाषा नहीं समझ आए, तो उसके घर की भाषा को, मातृभाषा को, विद्यालय की भाषा से जोड़ने की कोशिश करनी चाहिए। ऐसा संभव नहीं हो पाता क्योंकि शिक्षक ऐसा करने की कोशिश ही नहीं करते, क्योंकि अभिभावक खुद अपने बच्चों पर ध्यान नहीं देते। वे लंबवत गतिशीलता पर बहुत ध्यान देते हैं इसलिए छात्रों के बीच दूरी बढ़ती रहती है और धनी एवं निर्धन में भी भाषा की वजह से दूरियां बढ़ती जाती हैं। आखिर में छात्र अंग्रेज़ी अथवा प्रदेश की सशक्त भाषा के प्रभाव में आ जाता है। इससे सबसे ज़्यादा नुकसान उसके घर की मातृभाषा का ही होता है जिसे बोलना ऊपर दिए कारणों से उसे एक शर्मनाक स्थिति में डाल देता है। इस कारण बहुत से ऐसे बच्चे मिल जाएंगे जो अपने घर की भाषा तो थोड़ा बहुत बोल लेते हैं लेकिन उसे पढ़, लिख नहीं सकते हैं।

यह कहना गलत नहीं होगा कि अन्य देशों की तरह जहां ब्रिटिश साम्राज्य था भारत में भी अंग्रेज़ी मिलजुल गई और एक-दूसरे प्रकार की अंग्रेज़ी के रूप में विकसित हो गई है। साथ ही यह भी कि बहुत ही कम लोग मानेंगे कि अंग्रेज़ी अब सिर्फ अंग्रेज़ों की भाषा रह गई है। इस प्रकार अंग्रेज़ी भाषा बहुत महत्वपूर्ण हो गई है। साथ ही यह भाषा कुछ लोगों के लिए वरदान तो बहुतों के लिए अभिशाप बनती जा रही है। यों तो भारत में प्रायः सभी लोग अंग्रेज़ी पढ़ते हैं, बोलते हैं, एवं लिख लेते हैं, परंतु प्रश्न उठता है कि वे लोग किस मानसिकता से उस विदेशी भाषा का उपयोग करते हैं, उनकी क्या मनोवृत्ति रहती है। क्या वे उसके प्रयोग में बहुत गर्व महसूस करते हैं या उसे प्रयुक्त करने में शर्माते हैं? इस प्रकार अंग्रेज़ी और उसकी संस्कृति हम भारतीयों के लिए एक शक्ति या एक अभिशाप का चिह्न बन कर रह गई है।

जब किसी भाषा को पढ़ाया जाता है, तो निश्चय ही उसकी संस्कृति अनायास ही पढ़ाई जाती है। यह अंग्रेज़ी के लिए उतना ही सही है जितना कि अन्य भारतीय भाषाओं के संदर्भ में। भाषा के साथ उसकी संस्कृति स्वतः संचारित हो जाती है, और इसमें ग्लानि नहीं होनी चाहिए। जहां भी द्वितीय भाषा या विदेशी भाषा पढ़ाई जाती है, उसका प्रभाव यह है कि लोग उस भाषा की संस्कृति, समाज तथा मूल्यों के बारे में जानना चाहते हैं। ऐसा करने में उस संस्कृति की तो सिर्फ जानकारी दी जाती है, उसे पालन करने का कोई निर्देश नहीं होता। किसी अन्य संस्कृति को जानने से खुद की संस्कृति तो खत्म नहीं हो जाती।

कुछ अन्य लोग इस दृष्टिकोण के हैं कि जो भाषा सिखाई जा रही है, उससे जुड़ी संस्कृति को विद्यार्थियों से दूर रखा जाए और किसी प्रकार का सांस्कृतिक आरोपण न हो क्योंकि कोई भी भाषा उस समुदाय की आत्मा होती है और उसकी मनोवृत्ति को दर्शाती है। कभी-कभी हम भाषा पढ़ाते हुए लोगों की विचारधारा, चिंतन को भी बदल देते हैं। यह बहुत हानिकारक है क्योंकि यह सांस्कृतिक

साम्राज्यवाद की श्रेणी में आ जाएगा। हमें भाषा-शिक्षण में इस सांस्कृतिक हस्तक्षेप से बचना होगा क्योंकि प्रत्येक भाषा अपने साथ संस्कृति का साहचर्य लिए हुए है।

भाषा किसी भी समुदाय की विचारधारा को बदल सकती है, खासकर जब वह राजनैतिक, आर्थिक क्षेत्र में प्रबल एवं मुख्य हो। इसलिए भाषा तथा उसकी संस्कृति, कभी-कभी शोषण का कारण बन सकती है। जैसे कि मानक भाषा के समक्ष, बोलियों या घरेलू भाषाओं की स्थिति देखी जा सकती है। इससे विद्यार्थियों में कई प्रकार की हीन भावनाएं आ जाती हैं।

हमारे देश में अंग्रेज़ी प्रत्येक प्रदेश में किसी न किसी स्तर पर अनिवार्य रूप में पढ़ाई जा रही है। इसलिए हमें विद्यार्थी को राज्य की संस्कृति तथा अंग्रेज़ी दोनों पढ़ाने होते हैं। साथ ही, भारतीय संस्कृति की शिक्षा देनी होती है। हम सभी संस्कृतियों का समान आदर करें, इसके अतिरिक्त हमें यह भी ध्यान देना होगा कि हम अंग्रेज़ी संस्कृति का उपहास न करें। अंग्रेज़ी सीखने में जो भी अड़चनें आएं, उनके बावजूद उसका सम्मान करना होगा।

इससे हम संस्कृति को सीखेंगे ही, साथ ही दूसरों के प्रति धैर्यपूर्ण व्यवहार करना भी सीख लेंगे। हमारा देश एक बहुसंस्कृतिप्रधान देश है। अतः यह उसके लिए विशेषकर ठीक होगा। यहां हर प्रकार के लोग रहते हैं, हर राज्य में रहते हैं और क्योंकि अंग्रेज़ी बहुप्रचलित भाषा है, सिखाई जाती है, किसी रूप में बोली जाती है, इस कारण भारत में अंग्रेज़ी भी बहुसंस्कृतीय हो जाती है... अंतर प्रांतीय स्तर पर भी। अन्य संस्कृतियों की एक संवेदनशील जानकारी, प्रशंसा और सम्मान, अंग्रेज़ी की कक्षा में राष्ट्रीय एकता को बढ़ावा देगी एवं विविधता में एकता के भाव को जीवंत करेगी।

निस्संदेह अंग्रेज़ी संसार की बहुत महत्वपूर्ण भाषा बन गई है। क्या मैं यह सुझाव देने का दुस्साहस कर सकता हूं जिसमें सब मिलकर एक ऐसी अंग्रेज़ी भाषा का निर्माण करें जो सभी जगह समझी जा सके, आसानी से पढ़ी जा सके, जिसमें भाषाई एवं सांस्कृतिक आतंकवाद तथा साम्राज्यवाद की बू तक न हो। हमें ऐस्पेरांटो को मन में रखना होगा।

एक विडंबना यह भी है कि विदेश में बात करने के लिए संयुक्त राष्ट्र में हम गर्व से हिंदी में बात कर सकते हैं। साथ ही स्वचालित यंत्र आधारित अनुवाद की सुविधा वहां तुरंत उपलब्ध है। परंतु भारत में, अपने ही देश में, अपने ही संसद सदन में एक-दूसरे से बात करने, बोलने और समझाने के लिए अंग्रेज़ी का इस्तेमाल ज़रूरी हो जाता है।

अंग्रेज़ी भाषा भारत के लिए एक बड़ा सांस्कृतिक मुद्दा भी है। इसका हमारे ऊपर बहुत असर पड़ा है इसलिए आजकल अंग्रेजी के प्रति कुछ प्रतिरोध है। हो सकता है ऐतिहासिक कारणों की वजह से, बिना अंग्रेज़ी के हम शायद इतनी सुविधाएं न पाते। वैसे तो अंतरराष्ट्रीय अंग्रेज़ी अपने ही तरीके से अपना प्रभाव जता रही है और हम भारतीय, आज के परिवेश में खुद अंग्रेज़ी-भाषा एवं संस्कृति की ओर आकर्षित हो रहे हैं। अंग्रेज़ी भाषा प्रदर्शन के लिए, रोग के लिए है। इसके साथ स्तर का, अहं का भाव भी है। इसलिए अंग्रेज़ी रोज़मर्रा के जीवन के लिए कम, दफ्तर एवं आकर्षण के लिए

ज़्यादा है। इसी कारण हम अपनी संस्कृति एवं भाषा को, अंग्रेज़ी भाषा तथा संस्कृति से दूर रखते हैं, दूर रखने की कोशिश करते हैं पर अनायास ही कुछ मिश्रण हो जाता है। हमें द्विसांस्कृतिक होने में कोई संकोच नहीं होना चाहिए, और हमें द्विभाषीय (जिसमें एक भाषा अंग्रेज़ी हो) रहना चाहिए।

जिस प्रकार लैटिन का प्रयोग तथा प्रभाव रोमन साम्राज्य में हो गया था, उसी प्रकार आज यह स्थिति अंग्रेज़ी की भी है। हो सकता है कुछ समय बाद इतिहास कुछ और ही वक्तव्य दे, अंग्रेज़ी के प्रति। इसलिए हमारे देश में शिक्षण-व्यवस्था को हमेशा सांस्कृतिक संतुलन बनाए रखना होगा।

हम इस संदर्भ में यदि बहुभाषीय भारत में मातृभाषाओं का स्थान निश्चय एवं गौरवशाली ढंग से सुनिश्चित कर पाएंगे, तभी हम कुछ हद तक भारतीय संस्कृति का अनुसरण कर पाएंगे। फिर भारत की विभिन्न भाषाओं के बीच क्या ताल-मेल हो, और एक भाषा-समुदाय दूसरे भाषा-समुदाय से कैसी प्रतिक्रिया करता है, यह भी विचार करना होगा। यूरोपीय देश ज़्यादातर एकभाषीय हैं, एशिया के लोग, खासकर भारत के लोग ऐसे नहीं हैं। हम शायद जन्म से ही बहुभाषीय होते हैं। संस्कृति के साथ, जब हम यह पहलू देखें, तो बात और साफ हो जाएगी। संपूर्ण भारत के प्राचीन इतिहास में कहीं भी, किसी भाषा के आधार पर झगड़ा हुआ हो, नहीं दिखता। फिर ये कैसे हो गया कि पिछले 50 सालों में इतने भाषिक झगड़े हुए हैं, मुख्यतः अंग्रेज़ी हटाने या प्रतिष्ठित करने के नाम पर।

जिस प्रकार हम संस्कृति का उल्लेख करते हैं, उसी प्रकार धार्मिक संस्कृति एवं परंपराओं का भी सम्मानीय ढंग से उल्लेख होना चाहिए। एक धर्म, दूसरे धर्म के विरुद्ध न बोले, किसी भी प्रकार की भड़काने वाली स्थिति उत्पन्न न करें। ऐसे में हम अतीत की संस्कृति का किसी हद तक सम्मान कर पाएंगे।

हम सभी यह कहते हैं कि भारत एक बहुसांस्कृतिक राष्ट्र है और उसकी महत्ता अनेकता में एकता के भाव से जुड़ी हुई है जो आजकल सिर्फ एक प्रचलित कथन ही रह गया है। साथ ही, यह भी ज़ोर देकर कहना चाहिए कि भारत एक बहुधार्मिक राष्ट्र भी है। धर्म और संस्कृति दोनों ही जीवन के तथ्य हैं और ये बने रहेंगे, जीवन के हर क्षेत्र में, चाहे वह निजी हो या सामाजिक। मुख्य बात यह है कि अपनी विरासत की संस्कृति को याद रखते हुए, धर्म का किसी प्रकार से राजनैतिक प्रयोग नहीं होना चाहिए।

शिक्षकों को भी ठीक ढंग से संस्कृति के बारे में बताना, सिखाना चाहिए। बहुत प्रकार के अंतर्सेवा पाठ्यक्रम होते हैं - विषय-वस्तु संदर्भ, शिक्षण-विधि, जनसंख्या-शिक्षा, मूल्य, आदि परंतु किसी भी समय हमें धर्मनिरपेक्ष मूल्यों तथा प्रादेशिक संस्कृतियों के बारे में पाठ्यक्रम निर्मित करने का निर्देश नहीं मिलता, सुझाव नहीं आता। इसलिए जिन मान्यताओं को लेकर शिक्षक नियुक्त होता है, उन्हीं के साथ वह रहता है, उसी को पढ़ाता, अपनी अज्ञानता को बढ़ाता रहता है। किसी भी तरह से उदार दृष्टिकोण में नहीं ढलना चाहता है। यदि हमें संस्कृति पर ध्यान देना है, तो विषय-वस्तु के साथ-साथ सभी स्तर के शिक्षकों को संस्कृतियों के बारे में जानकारी देनी चाहिए – सीमित रूप में परंतु सही-सही।

शिक्षकों और शिक्षा-शास्त्रियों को इन मुद्दों पर अपने विचारों में परिवर्तन या संशोधन करना पड़ सकता है – जाति-प्रजाति तथा दलित आदि को लेकर, ताकि छात्रों का सही मार्गदर्शन हो सके। यह समस्या पूर्वोत्तर क्षेत्र के छात्रों के लिए बहुत जटिल हो सकती है। गंभीर चिंतन से ही ठोस समाधान हो सकता है, सिर्फ औपचारिक भाषणों से नहीं।

हमारी शिक्षा-व्यवस्था के अंतर्गत यदि विद्यालय दिवस या किसी अवसर पर कोई नाटक पेश करना हो, यदि वह अंग्रेज़ी में हो, तो उसे गरिमापूर्ण माना जाता है। भारतीय भाषाओं को उतना महत्व नहीं दिया जाता है। इसमें विद्यार्थियों को ठीक से तैयार करना शिक्षक का ही दायित्व होता है। यहां भी शिक्षक एक प्रकार का सांस्कृतिक दुस्साहस करते हैं। न भाषा हमारी, न छात्रों की, फिर भी उसी में अभ्यास, उसी में प्रदर्शन। सिर्फ इसलिए कि अंग्रेज़ी भाषा के अधिकाधिक व्यवहार से हमारी संप्रेषण दक्षता बढ़ेगी। यदि हिंदी या अन्य भारतीय भाषा में नाटक का मंचन करते तो शायद प्रभाव अधिक पड़ता। नाट्य मंचन के लिए सही बोलने की क्षमता एवं ढंग, नेत्र संचालन एवं भाव, स्पर्श, आत्मविश्वास, दृष्टिकोण इत्यादि आवश्यक होते हैं। इसलिए यदि हम उतना ही समय एवं ऊर्जा अपनी भाषा में लगाएं, तो हमें दोहरी सफलता मिल सकती है – नाटक मंचन में और भाषा के ग्रहण में।

आजकल बहुसस्कृतीय शिक्षा एक महत्वपूर्ण शैक्षणिक बिंदु हो गया है। भारत में इसकी महत्ता और अधिक है क्योंकि विदेशी संस्कृति के साथ भारतीय प्रांतीय संस्कृति का ज्ञान आवश्यक हो जाता है। सामाजिक अवस्था में इस गति से बदलाव आ रहा है कि हमारी शिक्षा-नीति में बहुसंस्कृतीय शिक्षा पर तुरंत ध्यान देना पड़ेगा। कोई भी प्रांत पूर्ण रूप से एकभाषिक एवं एकसांस्कृतिक नहीं है इसलिए शिक्षकों के लिए बी.एड.पाठ्यक्रम में इस दिशा में सुधार अत्यंत आवश्यक है। शिक्षक के विश्वास और धारणाएं उसकी कक्षा-शिक्षण को उतना ही प्रभावित करती हैं जितना कि कक्षा के बाहर शिक्षक का व्यवहार उसकी संस्कृति एवं संस्कार को दर्शाता है। इस प्रकार बहुसंस्कृतीय शिक्षा शिक्षक की खुद की जागरूकता से शुरू होती है। अनौपचारिक बातचीत से पता लगता है कि हमने उन्हें बहुसंस्कृतीय शिक्षा के बारे में चिंतन का कोई अवसर ही नहीं दिया। औपचारिक रूप से ही एक दूसरे के तीज त्यौहार मनाने को महत्व देने से कम से कम संस्कृति के बारे में सही जानकारी तो अधिकृत रूप से दे देनी चाहिए। अभी तो हम सिर्फ अधपके रूप से एक-दूसरे की संस्कृति, धर्म के बारे में ज्ञान एवं धारणाएं रखते हैं, जिस तरह कि भारतीय संस्कृति के बारे में हमारा ज्ञान अधूरा है, अधपका है। हमें बहुसंस्कृतीय शिक्षा के लक्ष्य निर्धारित करने चाहिए। यह शिक्षा किस प्रकार दी जाए, उसकी रणनीति सोचनी होगी तथा शिक्षकों को इसमें पूरा योग देने के लिए आकर्षित करना पड़ेगा।

इस प्रकार बहुसंस्कृतीय शिक्षा में उन धर्मों पर विशेष ध्यान देना होगा जो अल्पसंख्या में हैं। अल्पसंख्यक, आरक्षित, अनुसूचित जाति/अनुसूचित जनजाति एक प्रकार से चिह्नित हो गए हैं। हमें

एक मानवीय संबंध पद्धति अपनानी होगी जिसमें धैर्य, एकता को बढ़ावा मिल सके, जिसमें सभी समुदायों में आपस में अच्छी समझ पैदा हो। मुख्य बात यह होनी चाहिए कि बहुसंस्कृतीय-शिक्षण से हर प्रकार की सामाजिक समानता तथा सांस्कृतिक विविधता सामने आनी चाहिए। सांस्कृतिक भिन्नता के प्रति सम्मान तथा उत्सुकता का भाव होना चाहिए जिससे जीवन में उत्सव की स्थिति पैदा होगी। यह भाव पाठ्यक्रम के द्वारा निरंतर विकसित होना चाहिए। इससे समाज को भी बहुसांस्कृतिकता के प्रति अपना उत्तरदायित्व समझ में आएगा।

संस्कृति के साथ, शिक्षा का एक अंग उन पहलुओं पर भी होना चाहिए जिससे आज का भारतीय समाज ग्रसित है – वर्ग-संघर्ष, जाति-प्रथा, लिंग-भेद, स्त्रियों की पीड़ा, दमन, शोषण इत्यादि।

इस लेख में संस्कृति के प्रकरण का विभिन्न पहलुओं से विश्लेषण किया गया है। शिक्षक एक प्रभावशाली व्यक्ति है। भारत जैसे विविधतापूर्ण देश में संस्कृति एवं बहुसंस्कृति, बहुधार्मिकता का मुद्दा आवश्यक है, जिसे शिक्षकगण सार्थक कर सकते हैं। यदि हमारे शिक्षक दृढ़ता और आत्मविश्वास के साथ कुछ विषम परिस्थितियों का सामना करते हैं, तो बधाई के पात्र हैं। हमें अपने संकीर्ण दृष्टिकोण को छोड़ना होगा और यथार्थवाद पर आधारित शिक्षा-नीति में प्राण डालने होंगे। इस तरह से, सूझ-बूझ के साथ परस्पर विश्वास बढ़ाने वाले उपायों पर चर्चा करने से स्पष्ट रूप से हमारी शिक्षा-नीति एवं हमारे समाज को जागरूक तथा सफल बनाने की प्रेरणा मिलेगी। संस्कृति, धर्म को छोटे-छोटे बिंदुओं में बांटकर, उसको न समझने से रोज़मर्रा का जीवन अस्त-व्यस्त हो जाता है। हमें इस दिशा में बहुत काम करना है। आशा है, इस लेख के कुछ सुझावों से शिक्षक लाभांवित होंगे। ❑

राष्ट्रीयता एवं भारतीय समाज

कमलेश कुमार चौधरी

प्रवक्ता, शिक्षा विभाग
एम.जे.पी. रुहेलखण्ड विश्वविद्यालय, बरेली

राष्ट्रीयता किसी भी राष्ट्र के नागरिकों की वह दृढ़ रागात्मक अनुभूति होती है, जिसके कारण राष्ट्र के नागरिक अपने को एक सूत्र में बंधा अनुभव करते हैं। राष्ट्र के नागरिकों में इस भावना का विकास अपने राष्ट्र की भाषा, संस्कृति, परंपरा एवं इतिहास आदि के प्रति निष्ठा के भाव के विकसित होने से होता है। राष्ट्रीयता की भावना से ओत-प्रोत होने के कारण इन अमूल्य धरोहरों के संरक्षण तथा राष्ट्र की रक्षा एवं उत्थान हेतु व्यक्ति अपना सर्वस्व होम करने के लिए तत्पर होता है। किंतु आज समाज में जनसामान्य द्वारा जिस प्रकार का व्यवहार परिलक्षित हो रहा है उससे लगता है कि शिक्षा द्वारा राष्ट्रीयता की भावना का विकास नहीं हो पा रहा है।

राष्ट्रीयता किसी भी राष्ट्र के नागरिकों की वह दृढ़ रागात्मक अनुभूति होती है, जिसके कारण राष्ट्र के सभी नागरिक अपने को एक अनुभव करते हैं। किसी भी राष्ट्र के नागरिकों में राष्ट्रीयता की भावना का विकास अपने राष्ट्र की भाषा, इतिहास, संस्कृति एवं परंपरा आदि के प्रति निष्ठा के विकसित होने से होता है। यही कारण है कि किसी भी व्यक्ति को अपने राष्ट्र की इन अमूल्य धरोहरों पर गर्व होता है। राष्ट्रीयता की भावना से ओत-प्रोत होने के कारण इन अमूल्य धरोहरों के संरक्षण तथा राष्ट्र की रक्षा एवं उत्थान हेतु व्यक्ति अपने दायित्वों के निर्वहन हेतु अपना सर्वस्व होम करने के लिए तत्पर होता है। जिस राष्ट्र के नागरिकों में अपने राष्ट्र की भाषा, इतिहास, संस्कृति एवं परंपरा इत्यादि के प्रति सम्मान का भाव जितना अधिक होगा, वहां के नागरिकों में राष्ट्रीयता की भावना भी उतनी ही अधिक सशक्त होकर उभरेगी। राष्ट्रीयता की भावना के कारण ही भौगोलिक दृष्टि से अलग-अलग क्षेत्रों में निवास करने, धार्मिक भिन्नताओं तथा सामाजिक एवं आर्थिक असमानताओं के बावजूद लोग एकता के सूत्र में बंधे रहते हैं और राष्ट्रहित में उपरोक्त विषमताओं को भुलाकर एकजुट होकर निरंतर कार्य करने हेतु उद्यत रहते हैं।

वर्तमान भारतीय समाज के कार्य-कलापों पर दृष्टिपात करने से ऐसा प्रतीत होता है कि जन-सामान्य में राष्ट्रीयता की भावना का विकास पर्याप्त मात्रा में नहीं हो पा रहा है। क्षुद्र स्वार्थों की पूर्ति हेतु क्षेत्रवाद, भाषावाद, जातिवाद, संप्रदायवाद जैसे विवादों को बढ़ावा देकर राष्ट्रीयता के भाव

को जाने-अनजाने कुंठित किया जा रहा है। फलतः अब राष्ट्र की अस्मिता पर भी प्रश्न चिह्न लगने लगा है। देश के ही नागरिकों द्वारा ऐसी परिस्थितियां उत्पन्न की जा रही हैं, जो कहीं से राष्ट्रहित में नहीं हैं। खालिस्तान की मांग, कश्मीर समस्या, पूर्वोत्तर प्रांतों की समस्याएं इत्यादि इसके ज्वलंत उदाहरण हैं। मिलों, सरकारी कार्यालयों एवं औद्योगिक प्रतिष्ठानों में हो रही हड़तालें इस बात का प्रमाण हैं कि आपस में विश्वास का संकट पैदा हो गया है। जाति, धर्म, लिंग, संप्रदाय, भाषा इत्यादि के आधार पर समाज में भेद-भाव व कटुता के वातावरण का कारण राष्ट्रीयता की भावना का अभाव ही है। कर्मचारियों की हड़तालों से जहां एक ओर राष्ट्रीय विकास में अवरोध उत्पन्न हो रहा है, वहीं विभिन्न जन समस्याओं को लेकर सामाजिक संगठनों तथा राजनीतिक दलों द्वारा आंदोलन के समय सरकारी सम्पत्ति को क्षति पहुंचाने की मानसिकता राष्ट्रीयता की भावना के अभाव का प्रतिबिंबन करती है। संभवतः इसका कारण भारतीय समाज का वर्तमान आधुनिक स्वरूप है जिसका विकास भौतिकतावादी दृष्टिकोण के फलस्वरूप हुआ है।

जिस समुदाय या समाज के लोग नए समझे जाने वाले मूल्यों, विचारों, दृष्टिकोणों, संरचनाओं व संगठनों को अपनाने का प्रयास करते हैं, उसे आधुनिक समाज कहा जाता है। आधुनिक समाज में प्रायः व्यक्ति के लिए यह आवश्यक हो जाता है कि उसे किसी विशेष धंधे या कार्य में एक उच्च स्तर की शिक्षा या प्रशिक्षण प्राप्त हो, जिससे वह किसी आधुनिक प्रतिष्ठान में नौकरी पा सके और राष्ट्र के आर्थिक विकास में अपना योगदान दे सके। इन प्रतिष्ठानों में नौकरी पाने के पश्चात प्रगति पाने हेतु कई आवश्यक गुणों को भी उसे स्वयं में विकसित करना पड़ता है, यथा–समय पर आना, कठोर परिश्रम करना, वैज्ञानिक ढंग से कार्य करना, स्वच्छता, ईमानदारी, दक्षता, नई वस्तुओं एवं परिस्थितियों तथा वैज्ञानिक क्रियाओं को समझने का दृष्टिकोण रखना, नियमों के प्रति आदर भाव रखना, आज्ञाकारिता, समझौता करने में कुशलता रखना, मधुर मानवीय संबंध रखना आदि। उसे विषम व नई परिस्थितियों में अपना सामंजस्य स्थापित करना आना चाहिए। कदम-कदम पर होने वाली प्रतियोगिताओं तथा निराशाओं को सहन करने की अपरिमित शक्ति भी उसमें होनी चाहिए।

किसी भी समाज के आधुनिक होने के लिए उसके सदस्यों में उपरोक्त गुणों का होना आवश्यक है। इन गुणों से युक्त होने पर ही उस समाज के लोग बुद्धिमत्तापूर्वक अंधविश्वासों, कुतर्कों तथा आधुनिक युग की रोगग्रस्त, कठोर आत्मा वाली तथा यांत्रिक सभ्यता के दोषों से अपने को परे रख सकते हैं तथा राष्ट्र को आर्थिक विकास के पथ पर ले जा सकते हैं। परंतु किसी दूसरे समाज या राष्ट्र के स्वरूप को पूर्णरूपेण आत्मसात करके आधुनिक बनने की चेष्टा करने वाला कोई भी समाज, कभी भी आधुनिक समाज के रूप में अपने को विकसित नहीं कर सकता क्योंकि प्रत्येक समाज की अपनी अलग भौगोलिक, सामाजिक, आर्थिक, राजनीतिक, सांस्कृतिक पृष्ठभूमि एवं विशिष्टता होती है। इन सबके आलोक में ही किसी समाज को आधुनिक समाज के रूप में परिवर्तित करने का प्रयास किया जाना चाहिए। यदि कोई समाज अपनी पृष्ठभूमि

एवं विशिष्टताओं की उपेक्षा कर दूसरे समाज का अंधानुकरण करके आधुनिक समाज के रूप में विकसित होने का प्रयास करता है, तो वह निःसंदेह अपने समाज के मूल्यों, मान्यताओं, परंपराओं एवं आदर्शों की उपेक्षा करेगा और इससे राष्ट्रीयता की भावना का क्षरण ही होगा।

आज के भारतीय समाज की यह विडंबना है कि हम पाश्चात्य देशों का अंधानुकरण करके अपने समाज को आधुनिक समाज के रूप में तब्दील करने के अविवेकपूर्ण प्रयास में संलिप्त हैं। फलतः देश के नागरिकों में आधुनिक समाज के नागरिकों हेतु वांछित गुणों के विकास की ओर पर्याप्त ध्यान देने के बजाय, पाश्चात्य देशों के आधुनिक स्वरूप का अंधानुकरण कर हम उनके जैसा बनने का प्रयास करते जा रहे हैं। यह स्थिति भारतीय समाज के लिए फलदायी नहीं है। भारतीय समाज की पृष्ठभूमि पाश्चात्य देशों से भिन्न है। अतः भारतीय समाज को आधुनिक समाज के रूप में परिवर्तित करने का प्रयास भारतीय समाज की पृष्ठभूमि को दृष्टिगत रखकर ही किया जाना वांछनीय है।

पाश्चात्य संस्कृति की तर्ज पर भोगवादी संस्कृति की चकाचौंध के कारण समाज में उन्मुक्त यौनाचार एवं नशाखोरी की बढ़ती प्रवृत्ति को भारतीय समाज के उपयुक्त नहीं कहा जा सकता। आधुनिक समाज की मानक भोग-विलास की वस्तुओं को येन-केन प्रकारेण जुटाने की होड़ में, उचित-अनुचित का विचार किए बिना संलिप्त होने की प्रवृत्ति का समर्थन भी नहीं किया जा सकता। परंतु भारतीय समाज में आज ऐसा ही हो रहा है। वर्तमान भारतीय समाज की स्थिति यह है कि व्यक्तिगत स्वार्थ सर्वोपरि हो गया प्रतीत होता है। हम केवल अधिकारों की बात करते हैं, कर्तव्यों की नहीं। समाज में व्यभिचार, अनाचार, भ्रष्टाचार, घूसखोरी आदि अपनी पराकाष्ठा पर पहुंचे प्रतीत होते हैं। आधुनिक समाज की महत्वपूर्ण विशेषताएं -- ईमानदारी, समय पर आना, कठोर परिश्रम, मधुर मानवीय संबंध, आज्ञाकारिता, उत्तरदायित्व का लोगों में अभाव है, जबकि प्राचीन भारतीय समाज में इनकी बहुलता थी।

अतः यह आवश्यक हो जाता है कि भारतीय समाज को आधुनिक समाज के रूप में विकसित करने का प्रयास राष्ट्र की अमूल्य धरोहर, भाई-चारे, समानता, स्वतंत्रता, न्याय, कर्तव्यपरायणता एवं ईमानदारी इत्यादि को दृष्टिगत रखते हुए किया जाए। जन सामान्य में इन सामाजिक गुणों के विकास एवं उनके दृष्टिकोण को विस्तृत बनाने में शिक्षा की महती भूमिका को अस्वीकार नहीं किया जा सकता। राष्ट्रीय विकास में शिक्षा की महती भूमिका को स्वीकार करते हुए कोठारी कमीशन ने कहा है कि "भारत के भाग्य का निर्माण उसके अध्ययन-कक्षों में हो रहा है। स्कूल से निकलने वाले विद्यार्थियों की योग्यता एवं संख्या पर ही राष्ट्रीय पुनर्निर्माण के इस महत्वपूर्ण कार्य की सफलता निर्भर करेगी।"

निःसंदेह शिक्षा वह साधन है, जो लोगों में वांछित गुणों का विकास कर उन्हें इस योग्य बनाने में सक्षम है, जिससे वे अपने दायित्वों का निर्वाह सुचारु रूप से कुशलतापूर्वक कर सकें। शिक्षा द्वारा ही व्यक्ति अपने राष्ट्र की अमूल्य धरोहरों यथा भाषा, संस्कृति, इतिहास, परंपरा आदि से परिचित होता है। फलतः इन अमूल्य धरोहरों के प्रति उसमें निष्ठा के भाव का विकास होता है, जो राष्ट्रीयता की भावना के

विकास का आधार बनते हैं, जिससे व्यक्ति अपने क्षुद्र स्वार्थों से ऊपर उठकर राष्ट्र हित में अपने दायित्वों के निर्वहन हेतु अपना सर्वस्व होम करने के लिए तत्पर होता है। परंतु आज समाज में लोगों द्वारा जिस प्रकार के व्यवहार का प्रकटीकरण हो रहा है, उससे ऐसा लगता है कि शिक्षा द्वारा लोगों में राष्ट्रीयता की भावना का विकास पर्याप्त मात्रा में नहीं हो पा रहा है। इसका आशय यह नहीं है कि राष्ट्रीयता के विकास के संदर्भ में शिक्षा ने अपनी सार्थकता खो दी है, अपितु यह स्थिति इस तथ्य की ओर संकेत करती है कि राष्ट्रीयता के विकास हेतु शिक्षा के पाठ्यक्रम में जिन तत्वों का संयोजन होना चाहिए, शायद उनका संयोजन पर्याप्त मात्रा में नहीं हुआ है।

स्वतंत्र भारत की स्थितियां पूर्व की स्थितियों से निःसंदेह भिन्न हैं। इन स्थितियों को दृष्टिगत रखते हुए यह आवश्यक है कि शिक्षा का आयोजन इस प्रकार से किया जाए, जिससे यह जहां एक ओर देश के आर्थिक विकास में सहायक हो, वहीं दूसरी ओर विद्यार्थियों में उच्चतम मानवीय मूल्यों का विकास भी करे। वर्तमान समय में भौतिकतावादी दृष्टिकोण एवं जनसंख्या वृद्धि के फलस्वरूप बढ़ती हुई आवश्यकताओं के कारण शिक्षा में वैज्ञानिक एवं व्यावसायिक विषयों का महत्व बढ़ता जा रहा है। ऐसे विषयों का आकर्षण कम होता जा रहा है, जिनका उच्चतम मानवीय गुणों के विकास से सीधा संबंध है। फलतः विद्यार्थियों में उच्चतम मानवीय मूल्यों का विकास पर्याप्त मात्रा में नहीं हो पा रहा है। यही कारण है कि लोगों में सामाजिकता एवं राष्ट्रीयता की भावना के विकास की अपेक्षा अपने लिए जीने की प्रवृत्ति बढ़ती जा रही है। एतद् निमित्त भले ही दोष विभिन्न संचार माध्यमों द्वारा इस समाज में पाश्चात्य सभ्यता और संस्कृति के बढ़ते हुए प्रभाव को दिया जाए, किंतु शिक्षा भी इसके लिए कम ज़िम्मेदार नहीं है। अतः आवश्यकता इस बात की है कि शिक्षा का आयोजन इस प्रकार से किया जाए, जिससे वह अपना संबंध मात्र उत्पादकता (आर्थिक विकास) से ही न जोड़े अपितु लोगों में ईमानदारी, सांप्रदायिक सहिष्णुता, कर्तव्य-निष्ठा, भाई-चारे, न्याय प्रियता एवं सामाजिकता जैसे गुणों का विकास भी करें। वर्तमान भारतीय समाज में इन गुणों की महती आवश्यकता है। इन गुणों के विकास द्वारा ही राष्ट्र की बहुलतावादी संस्कृति अक्षुण्ण रह सकेगी और लोगों के जीवन में जाति, धर्म, सांप्रदायिकता, क्षेत्रीयता जैसी संकीर्ण भावना के स्थान पर राष्ट्रीयता की निर्मल धारा प्रवाहित हो सकेगी। इससे किसी भी क्षेत्र में कार्य करते समय लोग राष्ट्र हित को सर्वोपरि रख सकेंगे। अतः एतद् निमित्त शिक्षा का स्वरूप क्या हो, यह एक विचारणीय प्रश्न है।

लोगों में राष्ट्रीयता जैसे उच्चतम मूल्य के विकास हेतु शिक्षा के आयोजन में उन तथ्यों का समावेश करना होगा, जिससे विद्यार्थी देश के गौरवपूर्ण इतिहास की उन घटनाओं से परिचित हो सकें, जो त्याग एवं बलिदान से युक्त हैं एवं राष्ट्रीयता की भावना की सर्वोत्तम मिसाल हैं। साथ ही वे भारत की उस महान संस्कृति को भी जान सकें, जिसमें भारतीय मनीषी भूमि को जन्मदात्री मां के समान पावन तथा सुखद मानकर उसके ऋण को स्वीकार करते हैं। 'जननी जन्मभूमिश्च स्वर्गादपि गरीयसी' तथा 'माता भूमिः पुत्रोऽहं पृथिव्याः' जैसे विचार

मातृभूमि के प्रति प्रेम के द्योतक हैं। इनका प्रस्फुटन शिक्षा द्वारा होना चाहिए। भारतीय मनीषियों ने स्तुति में नदियों, पर्वतों, भू-प्रदेशों (तीर्थ के रूप में) तथा सप्तपुरियों को स्थान देकर मातृभूमि के प्रति अपनी कृतज्ञता ज्ञापित की है। इन भावों से विद्यार्थियों को अवगत कराकर, उनमें देश प्रेम की भावना जगाना शिक्षा का उद्देश्य होना चाहिए। भारतीय संविधान में समतामूलक समाज की स्थापना की बात कही गई है। समानता, स्वतंत्रता, भ्रातृत्व और न्याय समतामूलक समाज की स्थापना के आधार-स्तंभ हैं। अतः शिक्षा में ऐसे तथ्यों एवं क्रियाकलापों का समावेश किया जाना चाहिए, जिससे विद्यार्थियों में इन मूल्यों के प्रति आस्था विकसित हो। इसके लिए यह आवश्यक है कि विद्यालयी जीवन में विद्यार्थियों के सम्मुख एक ऐसा वातावरण उपस्थित हो, जहां जाति, धर्म, लिंग, क्षेत्र तथा अन्य किसी भी विषमता के आधार पर उनको हीन भावना का शिकार न होना पड़े। ऐसा होने पर ही इन मूल्यों के प्रति विद्यार्थियों में आस्था विकसित होगी और ये मूल्य संस्कृति का अंग बनकर पीढ़ी-दर-पीढ़ी हस्तांतरित हो सकेंगे। इससे एक ऐसे आधुनिक समाज का निर्माण होगा, जिसमें राष्ट्रीयता की अविरल धारा प्रवाहित हो सकेगी। परिणामतः भौगोलिक, भाषाई, धार्मिक तथा आर्थिक असमानता के बावजूद लोग एकता के सूत्र में बंधकर राष्ट्रीयता की भावना से युक्त हो, राष्ट्र के निर्माण में अपनी महती भूमिका के निर्वहन हेतु प्रस्तुत हो सकेंगे। ❑

संदर्भ

1. आटवे, ए.के.सी. (अनुवाद) शर्मा, ब्रजभूषण (1972) : ***शिक्षा और समाज,*** उत्तर प्रदेश हिंदी ग्रंथ अकादमी, लखनऊ।
2. बघेला, एच.एस, माहेश्वरी, पी.एन. और भोजक, बी.एल. (1985) : ***शिक्षा तथा भारतीय समाज,*** हर प्रसाद भार्गव, आगरा।
3. ब्राउन, एफ.जे. (1949) : ***एजूकेशनल सोशियोलॉजी,*** प्रेंटिस हाल, न्यूयार्क।
4. माथुर, एस.एस. (1985) : ***ए सोशियोलॉजिकल एप्रोच टु इण्डियन एजूकेशन,*** विनोद पुस्तक मंदिर, आगरा।
5. पाण्डेय, आर.एस (सं.) (1990) : ***भारतीय शिक्षा के विभिन्न आयाम,*** विनोद पुस्तक मंदिर, आगरा।
6. रुहेला, एस.पी. (1992) : ***भारतीय शिक्षा का समाज शास्त्र,*** राजस्थान हिंदी ग्रंथ अकादमी, जयपुर।
7. शुक्ला, के. और कुमार, के. (सं.) (1985) : ***सोशियोलॉजिकल प्रॉस्पेक्टिव इन एजूकेशन,*** चाणक्य पब्लिकेशन, दिल्ली।

अन्य भाषा में संप्रेषण की युक्तियां : हिंदी भाषा के संदर्भ में

विजयराघव रेड्डी

अध्यक्ष, आंध्र प्रदेश हिंदी अकादमी
हैदराबाद 500 007

इस लेख में पहले अन्य भाषा अधिगम में अंतर भाषा, अधिगम मनोविज्ञान का शिक्षण परिदृश्य, सीखने का परिदृश्य, अधिगम की युक्तियां, स्तब्धीकरण-प्रक्रिया पर व्यतिरेकी भाषा वैज्ञानिकों के दृष्टिकोण को संक्षेप में प्रस्तुत किया गया है। बाद में 'युक्ति' का मनोवैज्ञानिक परिप्रेक्ष्य प्रस्तुत करते हुए, अधिगंता की विविध युक्तियों के संदर्भ में उसकी अन्य भाषा के संप्रेषण में प्रयुक्त दोनों प्रकार की युक्तियों—बचाव की युक्तियों एवं संप्राप्ति की युक्तियों पर सोदाहरण विवरण प्रस्तुत करने का प्रयास किया गया है। ये सारे उदाहरण अहिंदी भाषी विशेषकर दक्षिण भारत के भाषा भाषियों की हिंदी की अभिव्यक्ति से एकत्रित किए गए हैं।

मनोविज्ञान के अधिगम अंतरण सिद्धांत को आधार बनाकर गत 40-45 वर्षों में विकसित व्यतिरेकी भाषा विज्ञान ने 'अंतर भाषा' की जो संकल्पना स्थापित की है, उसकी गठन-प्रक्रिया तथा उसके विकास क्रम एवं उसके विश्लेषण पर जो विशद् व्याख्याएं प्रस्तुत की गई हैं, वे वास्तव में मनोविज्ञान के कार्य क्षेत्र से सीधे संबंध रखती हैं। मनोविज्ञान को विशेष कर अन्य भाषा अधिगम मनोविज्ञान को यह भाषा विज्ञान की अनुपम देन है। भाषा विज्ञान के इस योगदान के प्रति मनोवैज्ञानिकों की क्या प्रतिक्रिया रही है, और इसका उन्होंने किस तरह से स्वागत किया, ये अलग प्रश्न है, जिनके उत्तर मनोवैज्ञानिक ही दे पाएंगे।

पारंपरिक व्यतिरेकी विश्लेषण में सीखने वालों की कठिनाइयों का पूर्वानुमान या त्रुटियों का जो पूर्वानुमान किया जाता था, उस पर जब प्रश्न चिह्न लगाए गए और जब उसकी पूर्वानुमानित क्षमता को संदेह की नज़र से देखा जाने लगा, तब व्यतिरेकी भाषा वैज्ञानिकों ने इस पर गहराई से विचार किया। तभी यह विचार समाने आया कि व्यतिरेकी विश्लेषण को यदि कारगर बनाना हो तो उसे स्रोत भाषा और लक्ष्य भाषा के व्यतिरेकी तत्वों के अतिरिक्त सीखने वाले की अधिगम प्रक्रियाओं जैसे तत्वों के विकासमान सिद्धांतों के ज्ञान का भी पता लगाना आवश्यक है। इससे व्यतिरेकी विश्लेषण विस्तृत एवं नवीन रूप में प्रकट हुआ। इस विस्तरण के परिणाम-स्वरूप अब व्यतिरेकी अध्ययन के अंतर्गत (क) अंतर भाषा, (ख) अधिगम की युक्तियां और (ग) स्तब्धीकरण की प्रक्रिया आदि पर विस्तार से विचार होने लगा है।

अंतर भाषा

अन्य भाषा अधिगम की प्रक्रिया विकासशील प्रक्रिया है। इस प्रक्रिया से गुज़रते हुए सीखने वाला अन्य भाषा को ग्रहण कर, दक्षता हासिल करता है। प्रक्रिया के किसी बिंदु पर यदि हम उसकी लक्ष्य भाषा को परखें और उसका विश्लेषण करें तो हम उसमें ऐसे कुछ तत्व पाएंगे जो स्रोत भाषा और लक्ष्य भाषा दोनों से संबंधित हैं। इस प्रकार के तत्वों से निहित सीखने वालों की विकासशील भाषा को व्यतिरेकी भाषा वैज्ञानिकों ने अंतर भाषा की संज्ञा दी है। इसी अंतर भाषा को अलग-अलग विद्वानों ने अलग-अलग नामों से अभिहित किया है। जैसे– निकटतम व्यवस्था, व्यक्तिनिष्ठ बोली, पथवाहक भाषा, संक्रमणिक भाषा, मध्यवर्ती भाषा आदि। इस अंतर भाषा के विश्लेषण में अधिगम प्रक्रिया पर विशेष महत्व देते हुए सीखने वाले की सीखी हुई भाषा को कई स्तरों पर अध्ययन करते हुए आगे के कार्यक्रम प्रशस्त किए जा रहे हैं।

अधिगम की युक्तियां

अधिगम की युक्तियों पर विचार करने से पहले अन्य भाषा अधिगम मनोविज्ञान पर दो बातें कहना आवश्यक प्रतीत होता है। इस मनोविज्ञान को दो परिदृश्यों के संदर्भ में देख सकते हैं। शिक्षण का परिदृश्य और सीखने या अधिगम का परिदृश्य। शिक्षण के परिदृश्य से अधिगम का परिदृश्य जुड़ा रहता है। शिक्षण का परिदृश्य उपागम, विधि और तकनीक पर मूलतः आधारित है। भाषा एवं उसके शिक्षण तथा अधिगम की प्रकृति से संबद्ध कुछ अन्योन्याश्रित मान्यताओं के समूहन को उपागम कहा जा सकता है। उपागम स्वतः सिद्धि है तो विधि प्रक्रिया है। उपागम के साथ विधि संबद्ध हो सकती है और विधि उपागम के साथ अविरोधक रहती है। जैसेः श्रवण उच्चारण उपागम, संज्ञानात्मक कूट उपागम, संप्रेषणपरक उपागम आदि। इन पर आधारित विधियां क्रमशः श्रवण-भाषण-विधि, प्रत्यक्ष विधि और संप्रेषणपरक शिक्षण विधि मानी जाती है। विधि वास्तव में शिक्षण सामग्री की प्रस्तुति से संबद्ध एक समग्र योजना होती है, इस तरह यह उपागम से संबद्ध युक्ति मात्र है, अर्थात उपागम से संबद्ध विचारों का शिक्षण में अनुप्रयोग है।

तकनीक विधि के साथ संगति रखती है और इस तरह उपागम के साथ समस्वरता बनती है। वास्तव में तकनीक कार्यान्वियन से संबंधित है जो कक्षा में शिक्षक के द्वारा घटित होती है। विविध प्रकार के ड्रिल, सांचा-अभ्यास, संवाद, वाक्य पूर्ति, किसी वार्तालाप में भूमिका निभाना आदि कुछ तकनीक हैं। यह विशिष्ट प्रकार की चाल या दांव-पेच या अधिष्करण या प्रयुक्ति है जो तात्कालिक लक्ष्य के संपादन के लिए काम में लाई जाती है। तकनीक अध्यापक की योग्यता/कार्यक्षमता/कला-कौशल/कृतित्व कक्षा की प्रकृति पर निर्भर रहती है। समस्या विशेष का निराकरण तकनीक विशेष से किया जाता है। अगर एक तकनीक किसी समस्या के समाधान के लिए कारगर साबित नहीं होती है, तो दूसरे तकनीक के चयन व प्रयोग की आवश्यकता होती है। इस तरह अन्य भाषा अधिगम

के शिक्षण परिदृश्य शिक्षक से संबंधित होते हैं। (एंथोनी एडवर्ड, 1922) लेकिन इसके विपरीत सीखने का परिदृश्य अधिगंता से संबंधित है।

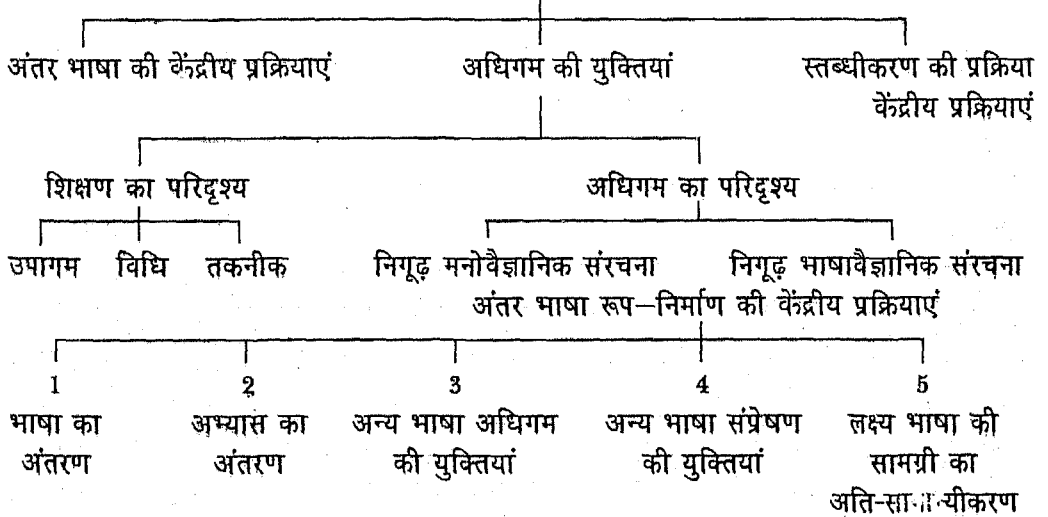
अन्य भाषा अधिगम को अर्थात सीखने वाला किस तरीके से भाषा सीखता है, इसे सिद्ध करना वास्तव में बड़ी समस्या है। सेलींकर (1975) कहते हैं कि अन्य भाषा सीखना वास्तव में एक अद्भुत मनोवैज्ञानिक प्रक्रिया है। लेकिन खेद है कि हम इस प्रक्रिया की प्रकृति से अनभिज्ञ ही हैं, इसका वैज्ञानिक विश्लेषण प्रस्तुत नहीं कर पाते। हम इतना मात्र जानते हैं कि इस प्रक्रिया में अन्य व्यावहारिक घटनाएं घटित होती हैं और इसके अनेक परिणाम भी निकलते हैं। वे आगे कहते हैं कि जब कभी सीखने वाला अन्य भाषा के वाक्यों का बोधन या निष्पादन करता है, तब उसकी मनोवैज्ञानिक संरचना पर जो प्रभाव पड़ता है उसे अधिगम प्रक्रिया पर घटित व्यावहारिक घटनाओं एवं व्यावहारिक परिणामों से जोड़ा जाता है। अन्य भाषा-अधिगम मनोविज्ञान में अगर हमारा लक्ष्य इस मनोवैज्ञानिक संरचना के प्रमुख पक्षों का उद्घाटन करना होता है तो तब हमारा मुख्य कार्य इसका पता लगाना होता है कि एकभाषा- भाषी द्विभाषी बनने की प्रक्रिया में कैसे अंतरभाषिकता से तादात्म्य स्थापित कर पाता है। अर्थात भाषाई संपर्क स्थिति में कैसे दो भाषाओं की ध्वनि, व्याकरण और शब्दार्थ प्रक्रियाओं से संबंध स्थापित कर लेता है।

लेन्नेबर्ग (1967) का उदाहरण देते हुए सेलींकर कहते हैं कि व्यक्ति के मस्तिष्क में एक ऐसी निगूढ़ मनोवैज्ञानिक संरचना रहती है जो अन्य भाषा सीखते समय निगूढ़ भाषा वैज्ञानिक संरचनाओं को अभिक्रियाशील करती है। दूसरे शब्दों में निगूढ़ मनोवैज्ञानिक संरचनाओं को सार्वभौम व्याकरण के जैविक प्रतिरूप तथा निगूढ़ भाषावैज्ञानिक संरचनाओं को सार्वभौम व्याकरण के भाषिक प्रतिरूप कह सकते हैं। अन्य भाषा अधिगम के अध्ययन में अधिगंता के स्रोत तथा लक्ष्य भाषा भाषियों की भाषा के संदर्भ में सीखनेवाले की अंतर भाषा की व्याख्या प्रस्तुत करनी चाहिए। स्रोत भाषा→अंतर भाषा→लक्ष्य भाषा– ये तीनों व्यावहारिक परिणाम हैं जो निगूढ़ भाषाई संरचनाओं से निष्पन्न हैं।

हमारा उद्देश्य मुख्य रूप से यहां उन मनोवैज्ञानिक प्रक्रियाओं का पता लगाना है जो अंतर भाषा व्यवहार को स्थापित करती है या अंतर भाषा गठन के कारण होती हैं। सेलींकर मानते हैं कि ऊपर उल्लिखित निगूढ़ मनोवैज्ञानिक संरचनाओं में पांच ऐसी केंद्रीय प्रक्रियाएं कार्य करती रहती हैं जो अन्य भाषा अधिगम में सहायक होती हैं और अंतर भाषा के रूप-निर्माण में योग देती हैं। वे केंद्रीय प्रक्रियाएं हैं–(1) भाषा का अंतरण (2) अभ्यास का अंतरण (3) अन्य भाषा अधिगम की युक्तियां (4) अन्य भाषा संप्रेषण की युक्तियां और (5) लक्ष्य भाषा की भाषाई सामग्री का अतिसामान्यीकरण। इनमें प्रथम तीन अधिगम के भाषा के निर्माणात्मक पहलू से संबंधित है तो अंतिम दोनों अधिगंता के भाषा के निष्पादन पहलू से संबंधित कहे जा सकते हैं। यह मोटे रूप से किया गया विभाजन मात्र है। अभ्यास की अंतरण प्रक्रिया आगे चल कर संप्रेषण की युक्ति का रूप धारण कर सकती है। इन पांच प्रक्रियाओं में से वर्तमान लेख का संबंध केवल चौथी प्रक्रिया,

संप्रेषण की युक्तियों से है, जिस पर आगे विवरण प्रस्तुत है।

स्तब्धीकरण प्रक्रिया

अन्य भाषा अधिगम के संदर्भ में अंतर भाषा में कुछ रचनाएं दृढ़ता से बनी रहती हैं। उन्हें दूर करने के उपाय करने पर, भले ही हमें यह लगे कि हमने उन्हें दूर किया है, फिर भी वे अंतर भाषा में स्थान बनाए हुए होती हैं। अक्सर वे तब स्थान प्राप्त कर लेती हैं जब सीखने वाला किसी जटिल और बौद्धिक विषय की ओर ध्यान ले जाता है या भाषा का प्रयोक्ता भावातिरेक की स्थिति में रहता है या व्यक्ति बहुत ही थकान या हलकेपन की अवस्था में रहता है। इस प्रकार की अंतर भाषा की संरचनाओं को व्यतिरेकी 'भाषावैज्ञानिक स्तब्धीकृत भाषाई तत्व कहते हैं। अन्य भाषा अधिगम में जो पांच केंद्रीय प्रक्रियाएं गिनाई गई हैं, उनके संदर्भ में भी हम स्तब्धीकृत तत्वों को देख सकते हैं। स्रोत भाषा के व्याघात के कारण से जो रचनाएं स्तब्धीकृत होती हैं, उन्हें भाषा अंतरण प्रक्रिया से; जिन स्तब्धीकृत रचनाओं को अभ्यासक्रम के संदर्भ में देखते हैं, उन्हें अभ्यास अंतरण प्रक्रिया से; स्तब्धीकृत रचनाएं अधिगंता के उपागम में शिक्षण सामग्री के संदर्भ में अभिज्ञेय हों तो, उन्हें अन्य भाषा अधिगम की युक्तियों से; स्तब्धीकृत तत्व अधिगंता के संप्रेषण के संदर्भ में अभिज्ञेय हों तो, उन्हें अन्य भाषा संप्रेषण की युक्तियों से; तथा स्तब्धीकृत तत्व लक्ष्य भाषा के नियमों के अतिसामान्यीकरण या अंतर भाषाई व्याघात के संदर्भ में देखते हैं तो, उन्हें लक्ष्य भाषा की भाषाई संरचनाओं की अतिसामान्यीकरण से संबद्ध कर सकते हैं। इस तरह स्तब्धीकरण प्रक्रिया का भी भाषा अधिगम की पांच केंद्रीय प्रक्रियाओं से संबंध स्थापित किया जा सकता है। प्रस्तुत विवरण का आरेख निम्न प्रकार से है :

युक्ति का मनोवैज्ञानिक परिप्रेक्ष्य

'युक्ति' शब्द पर विचार करते हुए सेलींकर कहते हैं कि मनोविज्ञान इसके उत्तर देने की स्थिति में नहीं है कि यह किन तत्वों से निर्मित होती है। अतः संप्रति युक्ति की व्यावहारिक परिभाषा देना संभव नहीं है। लक्ष्य भाषा में भाषा सामर्थ्य प्राप्त करते समय अर्थात उस भाषा पर दक्षता हासिल करते समय तथा उसमें संप्रेषण करते समय अधिगंता किन युक्तियों को काम में लाता होगा, इसकी जानकारी भी हमारे पास उपलब्ध नहीं है। कुछ लोग युक्तियों को संस्कृति सापेक्ष मानते हैं। वे कहते हैं कि यह कुछ पारंपरिक संस्कृतियों में आलापन को सीखने का एक साधन मानी जाती रही है। जैसे चीन की परंपरा में कंठस्थ करना और सस्वर पाठ का बड़ा महत्व है। पाठ को कंठस्थ करना और उस का सस्वर पाठ करना किसी भी विषय पर दक्षता हासिल करने के लिए चीनी शिक्षा व्यवस्था में आवश्यक माना गया है। अंग्रेज़ी सीखने वाले चीनी छात्रों ने इस बात को स्वीकार किया है कि शब्दावली एवं समस्यामूलक संरचनाओं को सीखने में तथा बोलना सीखने के लिए कंठस्थ करना तथा सस्वर पाठ करना अधिक कारगर साबित हुआ है। (हुआङ, 1987)

कंठस्थ करने के लिए सीखने वाला कुछ स्मृति सहायक तकनीकों को भी अपनाता है। लेकिन ये तकनीक प्रायः वैयक्तिक होती हैं। कंठस्थ करने की प्रक्रिया को दो रूपों में विभाजित कर सकते हैं– यांत्रिक रूप से कंठस्थ कर लेना या रटना और सार्थक रूप से कंठस्थ कर लेना। इन में से पहले को भाषाई रूपों को सीखने में और दूसरे को भाषा प्रयोग या प्रकार्य को सीखने में उपयोगी समझते हुए आजकल इस पर शोध कार्य भी किए जा रहे हैं। स्मृति सहायक तकनीक तथा कंठस्थ करने की प्रक्रिया को टेरोन (1985) अधिगम की युक्तियों में सम्मिलित करना चाहती हैं, न कि संप्रेषण की युक्तियों में।

कुछ भी हो, इतना तो सत्य है कि सीखने वाला चाहे वह चेतनावस्था में हो या उप चेतनावस्था में, जब यह महसूस कर लेता है कि उस के पास लक्ष्य भाषा के कुछ पक्षों के सामर्थ्य का अभाव है तब वह कुछ युक्तियों की शरण में अवश्य आ जाता है। लेकिन वस्तुतः ये युक्तियां क्या-क्या होती हैं और उनकी सहायता से कितनी कामयाबी वह हासिल कर पाता है, ये सब अनुमेय ही है। वुगेइस्की (1970) कहते हैं कि द्वितीय विश्व युद्ध के पश्चात के भाषा अधिगम पर हुए शोध कार्य ने यह उद्घाटित किया कि प्रत्येक अधिगंता एक दूसरे से भिन्न पृष्ठभूमि लिए आता है और वह मात्र अनुकरण व अभ्यासों की पुनरावृत्ति से नहीं सीखता। भले ही एक ही शिक्षण सामग्री का प्रयोग किया जाता हो, अधिगंताओं के सीखने के उपागम अपने-अपने अर्थात निजी व वैयक्तिक होते हैं।

अधिगंता अपने विशिष्ट तौर-तरीके से सीखता है। यह विशिष्ट निजी व वैयक्तिक तौर तरीके ही दूसरे अर्थ में अधिगम की युक्तियां हैं। युक्तियों में व्यक्ति सापेक्षता के साथ क्रमबद्ध व्यवस्था भी रहती है। अतः इन में व्यक्ति सापेक्ष संयोजन रहता है। पूर्वानुभव के आधार पर पहले जो युक्तियां सफलता से काम में लाई गई हैं, उनके आधार पर

भिन्न-भिन्न युक्तियों को अधिगंता काम में लाता है।

युक्ति, प्रक्रिया, योजना, उपाय, दांव-पेच और चाल आदि अधिगंता के अधिगम के संदर्भ में समानार्थक हैं या भिन्न संकल्पनाओं के द्योतक हैं? अगर भिन्न हैं तो किस तरीके से भिन्न हैं? इन प्रश्नों के उत्तर देने के कुछ प्रयास भाषा वैज्ञानिकों ने किए हैं। लेकिन इन में एकमत स्थापित नहीं हो पाया। ये ऐसे सूक्ष्म व व्यक्तिनिष्ठ मनोवैज्ञानिक क्रियाकलाप हैं, जिनकी सार्वभौम व्याख्या प्रस्तुत करते हुए एक दूसरे के अंतर को स्पष्ट करना कदाचित संभव ही नहीं है। फिर भी युक्ति एवं प्रक्रिया पर उपलब्ध विचारों में से कुछ यहां प्रस्तुत हैं।

युक्ति और प्रक्रिया को कुछ लोग समान मानते हुए कहते हैं कि इन दोनों में सरलीकरण की स्थिति रहती है। कुछ के विचार में युक्ति समस्या विशेष के समाधान से संबंधित है, अत: यह सामयिक है। लेकिन प्रक्रिया व्यवस्थीकृत उपायों की शृंखला होती है, जो अतिकालिक होती है। कुछ लोग युक्ति को प्रक्रिया के अंतर्गत उपवर्ग मानते हैं। कुछ लोग इन दोनों को एक ही मानसिक क्रिया से संबद्ध मानते हैं। बियालस्टाक (1987) के अनुसार युक्ति ऐच्छिक और प्रक्रिया अनिवार्य मानसिक क्रियाओं से संबद्ध है। इसी प्रकार कुछ लोग प्रक्रिया को सार्वभौमिक और युक्तियों को वैयक्तिक मानते हैं। कुछ के मतानुसार प्रक्रिया का संबंध चेतनावस्था से है, और युक्ति का संबंध चेतनावस्था से नहीं है।

जिस प्रकार से युक्ति, प्रक्रिया, योजना आदि की सुनिश्चित व्याख्या करना मुश्किल है, उसी प्रकार युक्तियों को अधिगम की युक्तियों एवं संप्रेषण की युक्तियों में विभाजित करना कठिन है। अक्सर सीखने वाला यह महसूस करता है कि उसकी सीखी हुई भाषाई संरचनाएं कुछ अवसरों पर अभिव्यक्ति करने के लिए पर्याप्त नहीं हो रही हैं। फिर भी उसे मजबूरन उस संदर्भ में अभिव्यक्ति करनी पड़ती है। अभिव्यक्ति को सफल बनाने की चेष्टा में उसे कुछ युक्तियों को काम में लाना पड़ता है। ये युक्तियां संप्रेषण की युक्तियां कहलाती हैं। (रेड्डी, 1986) अन्य भाषा सीखने वाला अपनी अंतर भाषा की संप्रेषण प्रक्रिया में जिन तौर-तरीकों को काम में लाता है, ये ही वास्तव में उसकी संप्रेषण युक्तियां मानी जाएंगी। कभी इनसे उसे सफलता भी प्राप्त हो सकती है। अधिगंता की संप्रेषण-युक्ति ने भाषाई संप्रेषण के उद्देश्य की पूर्ति की है कि नहीं, यह इस बात पर भी निर्भर करता है कि लक्ष्य भाषा के मातृभाषियों की सहृय-सीमा कितनी है। अंतर भाषा में संप्रेषण की युक्तियां, सीखने वाला जिन समस्याओं का सामना करता है, उनसे संबद्ध रहती हैं। इन समस्याओं का समाधान ढूंढ़ते हुए अधिगंता सचेतन में योजना बनाता है। इस योजना के अंग होते हैं, संप्रेषण की युक्तियां। कैंडलिन (1984) कहते हैं कि अंतर भाषा के संप्रेषण को दो दृष्टियों से देख सकते हैं :

(1) अधिगंता का समस्या रहित प्रयोग। चाहे ये प्रयोग देशीय भाषा-भाषी के अनुरूप हो या विशिष्ट हो या त्रुटिपूर्ण यह अलग बात है।
(2) समस्यामूलक प्रयोग। यहां समस्याओं को सुलझाने के लिए अधिगंता कुछ युक्तियों का प्रयोग करता है। ये ही संप्रेषण की युक्तियां हैं।

स्कोलफ़ील्ड (1987) ने ठीक ही कहा है कि भाषा की अभिव्यक्ति के दौरान उठने वाली समस्याओं के समाधान के रूप में प्रकट होने वाली संप्रेषण की युक्तियों पर इन दिनों अधिक विद्वानों का ध्यान गया है। वे आगे कहते हैं कि इन युक्तियों के संबंध में शोध करके इन पर विशद् व्याख्या प्रस्तुत करना आसान नहीं है। इन युक्तियों का पता लगाने के लिए इन दिनों चित्रों के माध्यम से पहले मातृभाषा में, बाद में लक्ष्य भाषा में कहानी कहलवाने की जो विधि काम में लाई जा रही है, उसकी भी अपनी सीमाएं हैं।

संप्रेषण के समय जब कोई समस्या उठ खड़ी होती है, तब उसके समाधान में सीखने वाले कैसे व्यवहार करते हैं? इस के उत्तर में यही कहा जा सकता है कि यह व्यवहार निजी व वैयक्तिक होता है। लेकिन गहराई से अवलोकन करने पर हमें यह प्रतीत होता है कि सामान्यतया अधिगंता समस्या के समाधान में दो उपागमों को प्रयोग में लाते हैं। एक है बचाव के व्यवहार से संबंधित और दूसरा है संप्राप्ति व्यवहार से संबंधित। पहले को बचाव की युक्ति और दूसरे को संप्राप्ति की युक्ति कह सकते हैं। पहले में संप्रेषण के लक्ष्य को ही बदला जाता है। दूसरे में संप्राप्ति पर आश्रित हो कर वैकल्पिक योजना के माध्यम से समस्या का समाधान करने की कोशिश की जाती है। इन दोनों युक्तियों को क्रमशः जोखिम बचाव की युक्ति और जोखिम स्वीकरण युक्ति भी कह सकते हैं।

कुछ विद्वान इन दोनों युक्तियों को बृहद युक्तियों के नाम से अभिहित कर इन के महत्व को स्वीकार करते हैं। फिर इनके अंतर्गत जो विविध प्रकार की उप युक्तियां गिनाई जाती हैं, उन्हें सूक्ष्म युक्तियों की श्रेणी में सम्मिलित करते हैं। (कार्डर, 1978)

प्रथम युक्ति की व्याख्या करते हुए राल्फु पामबर्ग (1983) कहते हैं कि इसके अंतर्गत संप्रेषण के लक्ष्य में जो बदलाव आता है, उसे विद्वानों ने विविध दृष्टिकोणों से परख कर विविध नाम दिए हैं। जैसे– प्रसंग परिहार (टैरोन), संदेश परित्याग (कोहेन), अभिप्राय प्रतिस्थापन (पारदी), अर्थविषयी परिहार (फाउन फोल्डर), संदेश लघुकरण (दुयास) आदि।

आगे कहा गया है कि जहां किसी विषय से संबंधित शब्दावली का ज्ञान अधिगंता को नहीं रहता हो, वहां वह उस प्रसंग का ही पूर्णतः परित्याग कर देता है। यह स्थिति प्रसंग परित्याग की होती है। जब लक्ष्य भाषा के उपयुक्त शब्द को याद करने में उसे कठिनाई होती है या उसे याद करने में वह असफल हो जाता हो, तब वह वाक्य के बीच में रुक कर वाक्य को अधूरा छोड़ देता है। इस स्थिति को संदेश परित्याग कहा जाता है। स्पष्ट और सही शब्द के ज्ञान एवं अभिव्यक्ति या मुहावरे के अभाव में जहां प्रयोक्ता अस्पष्ट या अति सामान्य अभिव्यक्ति का प्रयोग करता हो, वहां की स्थितियां अभिप्राय प्रतिस्थापन या अर्थविषयी परिहार या संदेश लघुकरण की मानी जा सकती हैं। ये स्थितियां जिन युक्तियों के कारण उत्पन्न होती हैं, उन सूक्ष्म युक्तियों को क्रमशः प्रसंग परिहार की युक्ति, संदेश परित्याग की युक्ति, अभिप्राय प्रतिस्थापन की युक्ति (या विषयी परिहार की युक्ति) और संदेश लघुकरण की युक्ति कहा जाता है।

प्रथम बृहद् युक्ति को कार्डर जोखिम बचाव की युक्ति कहने के बजाय संदेश समायोजन युक्ति कहना पसंद करते हैं। इस युक्ति के तहत अधिगंता जो उप या सूक्ष्म युक्तियों को काम में लाता है, उन्हें चार श्रेणियों में विभाजित करते हुए वे कहते हैं कि प्रसंग परिहार पराकोटि से संबंधित युक्ति है। जब प्रयोक्ता यह महसूस करता है कि उसके पास अभिव्यक्ति के लिए भाषाई तत्वों का पूर्णतः अभाव है, तब वह उस प्रसंग में प्रवेश ही नहीं करता।

प्रसंग परिहार से कम तीव्रतर स्थिति होती है, संदेश परित्याग की स्थिति। इस स्थिति में प्रयोक्ता अभिव्यक्ति करने का प्रयत्न तो शुरू कर देता है, लेकिन भाषाई तत्वों के अभाव में प्रयत्न को छोड़ देता है।

दूसरी स्थिति से (संदेश परित्याग से) कम तीव्रतर स्थिति होती है अर्थविषयी परिहार की स्थिति। इस स्थिति में प्रयोक्ता वास्तव में जो कहना चाहता है, वह उसे अभिव्यक्त तो नहीं कर पाता, लेकिन ईप्सित कथ्य से कुछ विचलित कथ्य की अभिव्यक्ति कर पाता है। फिर भी यह कथ्य मोटे रूप से थोड़ा बहुत जुड़ा रहता है।

संदेश समायोजन युक्ति के तहत अतीव न्यूनतम तीव्रतर की स्थिति होती है संदेश लघुकरण की स्थिति। इसमें प्रयोक्ता वास्तव में जो कहना चाहता है, उसे समग्रता में कह नहीं पाता, या प्रभावी ढंग से कह नहीं पाता, लेकिन कम शब्दों में या कम स्पष्टता से कहता है।

संप्रेषण की समस्याओं के समाधान में अधिगंता अपने संप्रेषण साधनों का विस्तरण कर लेता है। ऐसी दशा में काम में लाई जाने वाली बृहद् युक्ति को जोखिम स्वीकरण या जोखिम संचालन युक्ति कहते हैं। इस बृहद् युक्ति को कार्डर संसाधन विकास युक्ति कहना पसंद करते हैं। इस बृहद् युक्ति के अंतर्गत कई सूक्ष्म युक्तियां काम में लाई जाती हैं। ये वास्तव में भाषा संप्राप्ति की युक्तियां होती हैं। ये युक्तियां भाषा के सभी स्तरों पर काम में लाई जाती हैं। किसी कमी की पूर्ति में ये युक्तियां प्रयुक्त की जाती हैं। अतः कुछ विद्वान इन्हें प्रतिपूरक युक्तियां भी कहते हैं।

जिस तरह से प्रथम बृहद् युक्ति के अंतर्गत सूक्ष्म युक्तियों की क्रमबद्ध व्याख्या की जा सकती है, उस प्रकार की क्रमबद्धता इन में नहीं देखी जाती। इस बृहद् युक्ति के अंतर्गत एक साथ एकाधिक सूक्ष्म या उप युक्तियों का भी प्रयोग देखा जाता है। ये सारी जोखिम भरी युक्तियां हैं, क्योंकि इन के कामयाब न होने पर संप्रेषण के समाप्त हो जाने और गलतफहमी पैदा होने का डर बना रहता है।

इन युक्तियों में सर्वविदित युक्ति 'उधार की युक्ति' मानी जाती है। लक्ष्य भाषा को छोड़ कर भाषाई संसाधन का प्रयोग करना ही उधार की युक्ति है, जिसमें अनुमान से ईज़ाद किए गए या दूसरी भाषा से उधार लिए गए भाषाई तत्व सम्मिलित होते हैं, जो कमोबेशी में लक्ष्य भाषा के निकट के तत्व होते हैं। ये भाषाई तत्व अधिगंता की अंतर भाषा की दक्षता के अनुरूप परिमित रहते हैं। उधार के तत्वों में अधिक जोखिम से भरी युक्ति होती है— 'भाषा परिवर्तन' युक्ति। अर्थात लक्ष्य भाषा के स्थान पर अपनी पूर्व ज्ञात दूसरी भाषा में अपनी बात कह डालना।

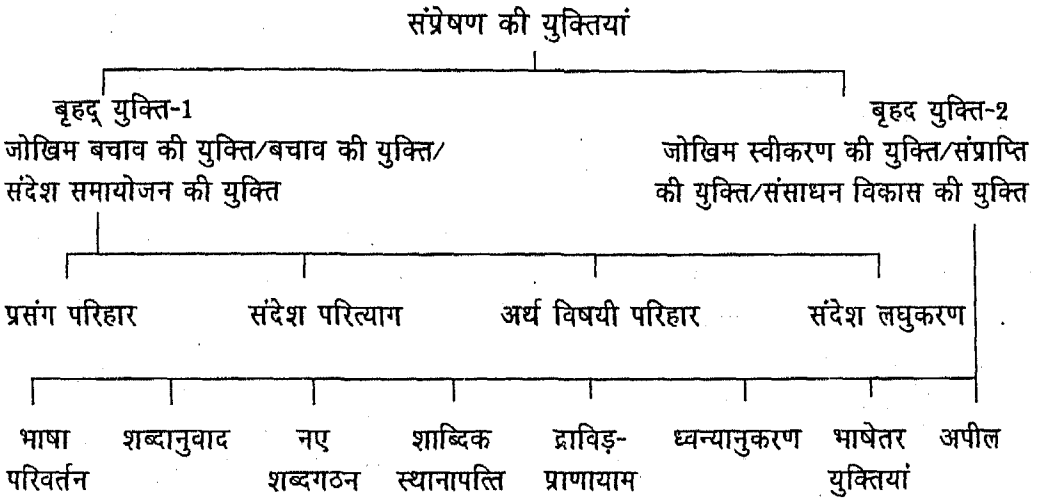
उधार की युक्ति से कम जोखिम भरी युक्ति होती है, पदान्वयन या वाक्जाल या द्राविड़ प्राणायाम की युक्ति अर्थात सीमित भाषाई दक्षता के कारण घुमा-फिरा कर समस्या का समाधान करना। इस प्रक्रिया में अभिव्यक्ति बेढंग तो लगती है, लेकिन कथ्य किसी न किसी तरह संप्रेषित किया जाता है।

पदान्वयन की तुलना में जोखिम से भरी युक्ति का प्रयोग वहां किया जाता है, जहां प्रयोक्ता भाषेतर संसाधनों के विकास की शरण में जाता है। जैसे : अवायिक संप्रेषण की तरफ झुकना या संकेतों से अपनी बात कहने का प्रयास करना या संवादांतर्गत व्यक्तियों से सही शब्द या संरचना के लिए अपील करना। कुछ विद्वान संवादांतर्गत व्यक्तियों से अपील को अलग युक्ति के रूप में सम्मिलित करते हैं। इन दिनों भाषेतर संसाधनों के विस्तार से संबंधित युक्तियों पर भी विस्तार से शोध किया जा रहा है। इस दूसरी बृहद् युक्ति के बारे में पामबर्ग (1987) ने टिप्पणी दी है कि अन्य भाषा अधिगंता संप्रेषण लक्ष्य को साधने के लिए अपनी समस्या का वैकल्पिक योजनाओं के द्वारा समाधान करता है। उनमें से अधिक उल्लेखनीय हैं :

- *भाषा परिवर्तन* : अर्थात अपनी मातृभाषा अथवा पहले सीखी हुई किसी दूसरी भाषा में अपनी बात कहना। इसे टॉरेन (1974) ने भाषा परिवर्तन और कार्डर (1978) ने कूट परिवर्तन कहा है तो बियालस्टाक (1980) ने देशीय भाषा के शब्दों का विदेशीकरण कहा है।
- *शब्दानुवाद* : अधिगंता कभी-कभी लक्ष्य भाषा की संरचना का शाब्दिक अनुवाद प्रस्तुत करता है। इस युक्ति में भले ही लक्ष्य भाषा के तत्वों का प्रयोग किया जाता हो, इन तत्वों की रचना का स्रोत, स्रोत भाषा का ज्ञान ही होता है। 'वारदी' के शब्दों में यह शब्दानुवाद की युक्ति है।
- *नए शब्दों का गठन* : लक्ष्य भाषा के उपर्युक्त शब्दों के ज्ञान के अभाव में अधिगंता कभी-कभी नए शब्दों का आविष्कार कर उनका प्रयोग कर देता है। 'वारदी' इसे शब्द गठन की युक्ति कहते हैं।
- *शाब्दिक स्थानापत्ति* : कभी अधिगंता ऐसे शब्दों का प्रयोग करता हैं जो लक्ष्य भाषा में सभी वांछित अर्थगत छायाओं का प्रतिपादन नहीं कर पाते। 'टोरेन' इसे शाब्दिक स्थानापत्ति युक्ति कहती हैं।
- *द्राविड़ प्राणायाम* : सही शब्द या रचना के ज्ञान के अभाव में सीखने वाला कभी-कभी अपनी बात को घुमा-फिराकर कहता है, कभी वर्णन करने लग जाता है और इस प्रकार वह वाक्जाल में फंसता है और ग्रहीताओं को भी फंसाता है। यह युक्ति पदान्वयन या वाक्जाल भी कहलाती है।
- *ध्वन्यानुकरण* : सीखने वाला कभी शब्द स्तर पर, और कभी पदबंध स्तर पर या वाक्य स्तर पर लक्ष्य भाषा के अनुरूप ध्वनियों का उच्चारण करने का प्रयास करता हैं। इस प्रयोग में कभी संयोग से संप्रेषण के लक्ष्य की सिद्धि हो जाती है। केस्यर ने इसे ध्वन्यानुकरण की युक्ति कहा है।
- *भाषेतर युक्तियां* : जब अधिगंता का भाषायी ज्ञान जवाब दे देता है तब वह भाषेतर युक्तियों

की शरण में जाता है। ऐसी स्थिति में वह इशारों, और मुख-मुद्राओं की नकल उतारते हुए संप्रेषण के लक्ष्य को साधने का प्रयास करता है।

- *अपील* : संप्रेषण के समय अपनी भाषाई दक्षता की कमी के कारण, कोई अन्य साधन न पाकर अधिगंता संवादांतर्गत व्यक्तियों से सही शब्द या रचना के लिए अपील करता है। हासद्रूब आदि इसे सहकारी युक्तियों में सम्मिलित करते हैं।

उपर्युक्त चर्चा के अनुसार अन्य भाषा के व्यवहार में अधिगंता के द्वारा प्रयोग में लाई जानेवाली संप्रेषण युक्तियों का विभाजन निम्नलिखित आरेख के द्वारा दिखाया जा सकता है :

इन दोनों बृहद् युक्तियों में से चयन या परिवर्तन न केवल बचाव व्यवहार या संप्राप्ति व्यवहार पर आश्रित है, अपितु समस्या की प्रकृति एवं तीव्रता पर भी आश्रित होता है। सीखने वाला एक ओर अपने बोझ को हल्का करने के उद्देश्य से अपने कार्य को सुगम बनाने की प्रक्रिया में तथा दूसरी ओर त्रुटियों से बचने के लिए बचाव की युक्ति का इस्तेमाल करता है। इस युक्ति का प्रयोग प्रायः प्रत्येक अधिगंता समय-समय पर कर लेता है। यह मानव स्वभाव का एक अभिलक्ष्य है।

कुछ विद्वान फरेह एवं केस्पर (1984) आदि बचाव की युक्ति को लघुकरण की युक्ति कहते हैं क्योंकि अनुमानित त्रुटि से बचने के लिए अधिगंता एक कठिन भाषाई तत्व को त्याग कर उसके स्थान पर किसी दूसरे तत्व का प्रयोग करता है जो उसको सरल प्रतीत होता है। अर्थात त्यक्त तत्व की तुलना में चुना गया तत्व लघुकृत होता है। लघुकरण की युक्ति रूपात्मक भी होती है और प्रकार्यगत भी। रूपात्मक लघुकरण का संबंध प्रकार्यगत लघुकरण अर्थात संप्राप्ति युक्ति से भी होता है।

इन दोनों को अंग्रेज़ी में 'फार्मल रिडक्शन स्ट्रेटजी' और 'फंक्शनल रिडक्शन स्ट्रेटजी' कहते हैं। भाषाई संरचना का कोई अधिगंता क्यों लघुकरण करता है और भाषा में कौन-सी ऐसी संरचनाएं हैं जिन के लघुकरण की संभावना होती है—ये दो ऐसे प्रश्न हैं जिनके उत्तर ढूंढने में व्यतिरेकी भाषा विज्ञान प्रयत्नशील है। जहां तक पहले प्रश्न का उत्तर है या तो अधिगंता त्रुटियों से बचना चाहता है या अपनी भाषा के प्रवाह को बढ़ाना चाहता है। इन दोनों के आधार पर रूपात्मक लघुकरण को हम त्रुटि से बचने के लिए घटित लघुकरण और संप्रेषण की सुगमता के लिए घटित लघुकरण के रूप में विभाजित कर सकते हैं। जहां तक दूसरे प्रश्न का संबंध है इस का कोई उत्तर नहीं दिया जा सकता। यह भाषा-भाषा में तथा अधिगंता-अधिगंता के अंतर पर निर्भर है।

वास्तव में त्रुटि-बचाव मनोवैज्ञानिक दृष्टि से निर्धारित है क्योंकि अक्सर शुरू-शुरू में अधिगंता का यह खयाल रहता है कि संप्रेषण में सफलता पाने के लिए भाषा संरचनाओं में सुधार की पूर्व आवश्यकता है। अधिगंता कुछ स्थानों पर यह समझता है कि कुछ रूपगत तत्वों के विलोपन से समाचार या अर्थ के संप्रेषण में कोई व्यवधान नहीं पड़ता और इससे भाषा प्रवाह में बढ़ाव होता है व संप्रेषण में सुगमता होती है। इस प्रकार के विलोपन को हेरोन (1984) उत्पादन युक्ति कहती हैं। वे आगे कहती हैं कि अधिगंता ऐसे स्थानों पर यह समझने लग जाता है कि भाषा रूप शुद्ध या व्याकरणिक न होने पर भी संदर्भ के अनुसार औचित्यपूर्ण हैं। अव्याकरणिक वाक्यों के प्रयोग से यदि कोई सीखने वाला अपने आशय का संप्रेषण कर पाता हो तो उसे सुधारने का प्रयास वह क्यों करना चाहेगा? मानक भाषा के प्रतिमान की आवश्यकता की बात को छोड़कर इस संबंध में विशेष बात यह है कि भाषा का संबंध बोलने वालों की अभिवृत्ति और सामाजिक रचना से जुड़ा रहता है। व्याकरणिकता या ध्वनि प्रक्रिया के प्रतिमानों से विचलित भाषा-प्रयोग वाले प्रयोक्ता से देशीय भाषा-भाषी दूरियां स्थापित करते हैं। भाषायी समाज से सफल निर्वाह के लिए व्याकरण सम्मत भाषा की भी आवश्यकता होती है। इस बात का खयाल अधिगंता कर नहीं पाता। रूपात्मक लघुकरण भाषा के सभी स्तरों पर होता है।

उपर्युक्त विवेचन के आधार पर अधिगंता के संप्रेषण की दो बृहद युक्तियों में से दूसरी बृहद् युक्ति के अंतर्गत जो आठ सूक्ष्म या उप युक्तियां मानी गई हैं, वे अन्य भाषा के अधिगंता व संप्रेषण में अधिक महत्व रखती हैं। लक्ष्य भाषा हिंदी के संदर्भ संप्रेषण की युक्तियों के कुछ उदाहरण प्रस्तुत हैं :

- **भाषा परिवर्तन/कूट परिवर्तन/उधार की युक्ति**

(क) मुख्य मंत्री का उपन्यास बहुत अच्छा है। (इसमें 'भाषण' शब्द के स्थान पर मातृभाषा के शब्द 'उपन्यास' का प्रयोग किया गया।)

(ख) मुझे बटर खरीदना है।
('मक्खन' शब्द के उच्चारण में कठिनाई के कारण प्रयोक्ता ने अंग्रेज़ी शब्द 'बटर' का प्रयोग किया है।)

(ग) अब 'बॉस' अच्छे 'मूड' में हैं। जाकर असलियत समझाइए। He will certainly consider your case. (शब्दों के साथ-साथ

वाक्य स्तर पर भी यहां कूट परिवर्तन किया गया है।)

- **शाब्दिक अनुवाद**

'मैं आते ही वह गया' मेरे आते ही वह गया के स्थान पर इस वाक्य का प्रयोग शाब्दिक अनुवाद का अच्छा उदाहरण है। तेलुगु और मराठी में वाक्य रचना इसी प्रकार होती है। जैसे तेलुगु में 'मेनू (मैं) वास्ताने वाडू वेलिलपोयाडू' अथवा मराठी में 'मी (मैं) वेताच तो गेला'।

- **नए शब्द गठन**

लक्ष्य भाषा के सही शब्द के ज्ञान के अभाव में अधिगंता अपनी अंतर भाषा में नए शब्दों का गठन कर उनका प्रयोग करता है। इस गठन में उसकी सृजनशीलता कार्य करती है। जैसे– गुब्बारे के लिए 'हवा की गेंद', रज़ाई के लिए 'ओढ़ने का बिस्तर', उपमा के लिए 'नमकीन हलवा' आदि इस युक्ति के नमूने हैं।

- **शाब्दिक स्थानापत्ति**

शाब्दिक स्थानापत्ति में अधिगंता (क) पूर्णतः नए शब्द की स्थानापत्ति करता है जो संदर्भ के अनुसार औचित्यपूर्ण नहीं होता, (ख) अर्थगत समीपवाले शब्द का प्रयोग करता है या (ग) बहिर-कोट शब्द का प्रयोग करता है जो संदर्भ के अनुकूल नहीं होता है। जैसे :

(क) टीले के लिए पहाड़ का प्रयोग।

(ख) टेबिल के लिए बेंच का तथा पलंग के लिए खाट का प्रयोग।

(ग) गाय या हाथी के लिए जानवर का प्रयोग।

- **द्राविड़ प्राणायाम**

'मुझे भूख लगी है, मैं सवेरे से भूखा हूं' वाक्यों की रचना न जानने या भूख/भूखा शब्दों के उच्चारण न कर पाने के कारण इन वाक्यों का प्रयोग 'मेरा पेट खाली है, सवेरे से मेरा पेट खाली है', इसके उदाहरण हैं।

- **ध्वन्यानुकरण**

'सदा सच बोलो, कभी झूठ मत बोलो' के लिए 'सदा सच बोल कबी जूट मत बोल' का प्रयोग ध्वन्यानुकरण का उदाहरण है।

- **भाषेतर युक्तियां**

अंग संचालन व मुख-मुद्राओं के ज़रिए भी अधिगंता अपनी बात को कहने की कोशिश करता है।

- **अपील**

अपील का एक उदाहरण इस प्रकार है :

प्रश्न : आप कॉलेज कैसे जाते हैं?

उत्तर : मैं कालेज उस .... उसे क्या कहते हैं? वह .... से ....

प्रश्न : वह कैसा दिखता है?

उत्तर : वह साइकिल नहीं ...... फट-फट (हंसी. .. नहीं वह साइकिल नहीं, मोटर भी नहीं .... साइकिल के साथ मोटर ....

प्रश्न : अच्छा मोटर साइकिल?

उत्तर : नहीं ... नहीं

प्रश्न : फिर मोपेड?

उत्तर : हां .... हां मोपेड, मैं मोपेड से कालेज जाता हूं।

उपर्युक्त आठ युक्तियों के अतिरिक्त कुछ विद्वान दो और युक्तियों को सम्मिलित करते हैं। वे हैं अति सामान्यीकरण और पुनः रचना।

अति सामान्यीकरण को सेलींकर आदि ने अंतर भाषा के निर्माण में कार्यरत पांच केंद्रीय प्रक्रियाओं में से एक माना है जिसे वे लक्ष्य भाषा की भाषाई सामग्री का अतिसामान्यीकरण कहते हैं। हिंदी के संदर्भ में इस युक्ति के कुछ उदाहरण प्रस्तुत हैं :

"इतना (हाथों से इशारा करते हुए) - अर्थात ज़्यादा खाइए, अपना घर समझिए। शर्म न कीजिए" के लिए यह प्रयोग –

"इतना-इतना खाता है। अपना घर समझता है। शर्म नहीं करता।"

पुनः रचना के अंतर्गत अधिगंता पहले एक रचना प्रयुक्त करता है। लेकिन उसे त्रुटिपूर्ण समझ कर उस के लिए पुनः रचना कर प्रयोग करता है। यह पुनः रचना सादृश्य पर आधारित हो सकती है। कभी-कभी यह पुनः रचना त्रुटिपूर्ण भी हो सकती है जबकि पहली रचना त्रुटि रहित होती है। जैसे–

मुझे दुःख है – मैं दुःखी हूं।
मुझे सुख है – मैं सुखी हूं।
मुझे भूख है – मैं भूखी हूं।

ऊपर के उदाहरण किसी सर्वेक्षण पर आधारित नहीं हैं। यह कहना भी मुश्किल है कि अमुक प्रयोग अमुक युक्ति के कारण प्रयुक्त है। यह भी संभव है कि अधिगंता एक साथ एकाधिक युक्तियों का भी प्रयोग करें। हिंदी के अन्य भाषा के रूप में सीखने वाले अहिंदी भाषियों के संप्रेषण में प्राप्त भाषा तत्वों के सही अध्ययन और विश्लेषण करने पर इन युक्तियों के सप्रमाण उदाहरण व इन पर विस्तृत व्याख्या प्रस्तुत की जा सकती है और इस व्याख्या के आधार पर शिक्षण में परिवर्तन लाए जा सकते हैं। ❑

संदर्भ

1. बुगेल्स्की, पी.आर. (1979) : ***प्रिंसिपल्स ऑफ लर्निंग एण्ड मेमोरी,*** प्रेजर पब्लिशिंग, न्यूयार्क।
2. कैण्डीन, सी.एन. (1984) : ***जनरल एडिशन स्ट्रेटजीज़ इन इंटर लैंग्वेज कम्यूनिकेशन,*** लांगमन, लंदन।
3. गुप्ता, मनोरमा (1984) : ***भाषा अधिगम,*** केंद्रीय हिंदी संस्थान, आगरा।
4. जेम्स, सी. (1981) : ***कंट्रास्टिव एनालिसिस,*** लांगमन, लंदन।
5. रेड्डी, विजयराघव (1986) : ***व्यतिरेकी भाषा विज्ञान,*** विन्दो पुस्तक मंदिर, आगरा।

शैक्षिक पर्यवेक्षण : व्यावहारिक परिप्रेक्ष्य

ज्ञानानंद प्रकाश श्रीवास्तव

प्रवाचक, क्षेत्रीय शिक्षा संस्थान
भोपाल

निरीक्षण से व्यापक पर्यवेक्षण केवल प्रशासनात्मक कार्यों को न देखते हुए, शैक्षिक क्रियाओं में योगदान करता है तथा शिक्षकों की शिक्षण दक्षता में निखार लाता है। समूह पर्यवेक्षण, विषयगत एवं संदर्भगत पर्यवेक्षण, सृजनात्मक पर्यवेक्षण आदि शैक्षिक पर्यवेक्षण के आधुनिक रूप हैं। पर्यवेक्षण प्रक्रिया के अंतर्गत योजना बनाना, अभिकल्प निर्मित करना, मार्गदर्शन करना, प्रेरित करना, लक्ष्य निर्धारण एवं वातावरणीय कारकों को नियंत्रित करना आदि सम्मिलित है। लेख के अनुसार त्रुटियां ढूंढ़ने से हटकर प्रोत्साहन, मार्गदर्शन, अभिप्रेरणा, पाठ्यक्रम निर्माण एवं सुधार, सेवाकालीन प्रशिक्षण एवं शैक्षिक मूल्यांकन आदि पर्यवेक्षक के मुख्य कार्य हैं।

पर्यवेक्षण का उद्देश्य शिक्षा संस्थाओं की शैक्षिक क्रियाओं में सहयोग प्रदान करना है। आधुनिक युग में पर्यवेक्षण केवल निरीक्षण मात्र नहीं है तथा इसका कार्य केवल प्रशासनात्मक प्रभुता का प्रयोग करना नहीं है बल्कि अध्यापकों की शिक्षण दक्षता में निखार लाना तथा शिक्षकों की समस्याओं के निदान में सहायता पहुंचाना है। यही कारण है कि विकसित देशों में आज समूह पर्यवेक्षण, शैक्षिक पर्यवेक्षण, विषयगत एवं संदर्भगत पर्यवेक्षण पर बल दिया जाता है। शैक्षिक पर्यवेक्षण का कार्य शिक्षण को उन्नत बनाना है, शैक्षिक कार्यक्रमों में परिवर्तन एवं परिवर्धन करना है। पर्यवेक्षण का संबंध छात्रों, सुविधाओं एवं संगठनात्मक कार्यों की अपेक्षा अध्यापकों तथा शिक्षण से अधिक है। साथ ही पर्यवेक्षण के अंतर्गत पाठ्यक्रम निर्माण एवं सुधार, सेवाकालीन प्रशिक्षण एवं शैक्षिक मूल्यांकन आदि भी सम्मिलित हैं। पर्यवेक्षण के मानव संबंध विचारधारा के अंतर्गत अध्यापकों का नैतिक साहस बढ़ाना, उनके साथ सहज संबंध स्थापित करना, उनकी आवश्यकताओं के प्रति संवेदनशील होना, उनके वैयक्तिक विकास में सहायता करना तथा उनके स्वतंत्र कार्यों के लिए वातावरण निर्मित करना भी पर्यवेक्षक के उत्तरदायित्व हैं। 'परिवर्तनशील-प्रक्रिया' के अनुयायी पर्यवेक्षण के कार्यों के अंतर्गत योजना बनाना, अभिकल्प (रूपरेखा) निर्मित करना, मार्गदर्शन करना, प्रेरित करना, लक्ष्य-निर्धारण तथा वातावरणीय कारकों को नियंत्रित करना आदि सम्मिलित करते हैं।

अतः शैक्षिक पर्यवेक्षण के अंतर्गत निम्नलिखित बिंदुओं पर ध्यान देना अपेक्षित है :

- प्रभावशाली पर्यवेक्षण सर्जनात्मक पर्यवेक्षण है।

पर्यवेक्षक का कर्तव्य है कि वह सहयोग का एवं मित्रवत वातावरण निर्मित करे। वह अध्यापकों के कार्यों में समन्वय स्थापित करे तथा अच्छे शिक्षण तथा कार्य में उसकी सहायता करे। वह अध्यापकों के सतत विकास एवं छात्रों की अभिवृद्धि को बढ़ावा दे।

- निरीक्षण एवं पर्यवेक्षण त्रुटियां ढूंढ़ने की प्रक्रिया नहीं है बल्कि सीखने-सिखाने की प्रक्रिया में उन्नति करना है। पर्यवेक्षण की पूरी प्रक्रिया प्रोत्साहन, मार्गदर्शन एवं अभिप्रेरणा से ओतप्रोत होनी चाहिए।
- अध्यापकों में आत्मविश्वास पैदा करना, कार्य के प्रति रुचि जाग्रत करना, उन्हें अभिप्रेरित करना, उनमें उत्साह भरना तथा कार्य के प्रति लगन पैदा करना अच्छे पर्यवेक्षण के लिए आवश्यक है।
- संदर्भगत पर्यवेक्षण इस बात पर बल देता है कि अध्यापकों में ईमानदारी, नैतिक निर्णय, न्यायशीलता एवं वैयक्तिक आदर के मूल्यों को जाग्रत करें।
- पर्यवेक्षक को अध्यापकों के विकास के स्तर को निर्धारित करना है जिससे उनमें अपेक्षित व्यावसायिक कौशलों का विकास हो। पर्यवेक्षक अध्यापकों के विद्यालयी क्रियाकलापों का अवलोकन करे, अध्यापकों से कक्षा के पूर्व एवं कक्षा के बाद विचार-विमर्श करे, उनकी क्रियाओं तथा शैक्षिक क्रियाओं पर प्रतिपुष्टि करे।
- पर्यवेक्षक लगातार इस बात को देखे एवं नियंत्रित करे कि अध्यापक कितने प्रभावशाली ढंग से शिक्षण कार्य करते हैं। वह अध्यापकों के साथ औपचारिक एवं अनौपचारिक बातचीत में विभिन्न शैक्षिक क्रियाकलापों पर चर्चा करे कि व्यावसायिक कौशलों में किस प्रकार निपुणता लाई जाए।
- ज़िला प्राथमिक शिक्षा परियोजना (डाइट) के अंतर्गत विद्यालय की प्रभावशीलता बढ़ाने के लिए शिक्षकों के प्राथमिक उन्मुखीकरण एवं प्रशिक्षण के उपरांत समय-समय पर विद्यालयों में एवं औपचारिक बैठकों में पर्यवेक्षक शैक्षिक गतिविधियों का पुनर्बलन करे एवं उनमें अपेक्षित परिवर्तनों के लिए बढ़ावा दे। डाइट राज्यों में ब्लाक शैक्षिक समन्वयक (बी आर सी) एवं संकुल शैक्षिक समन्वयक (सी आर सी) विद्यालयों के पर्यवेक्षण का कार्य कर सकते हैं। ज़िला शिक्षा एवं प्रशिक्षण संस्थान, ब्लाक संसाधन केंद्र एवं संकुल संसाधन केंद्र समन्वय, शैक्षिक सहायता एवं नियंत्रण का कार्य ले सकते हैं। इसके अंतर्गत तीन मुख्य कार्य हैं जैसे अध्यापकों एवं प्रशिक्षकों को प्रशिक्षित करने के कार्यक्रमों की रूपरेखा तैयार करना, ब्लाक तथा संकुल केंद्रों के लिए भौतिक साधनों का निर्माण तथा न्यूनतम अधिगम स्तर की प्राप्ति के लिए पाठ्यक्रम एवं कक्षा की गतिविधियों के लिए ऐसे कार्यक्रमों का आयोजन करना जिनमें छात्रों को सर्जनात्मक तथा बाल-केंद्रित क्रियाओं के अवलोकन एवं अभ्यास का अवसर मिले। संकुल संसाधन शिक्षक का इस प्रकार उन्मुखीकरण, प्रशिक्षण एवं सहायता की जाए कि वह अध्यापकों की गोष्ठी द्वारा, विद्यालय

में जाकर तथा अन्य अवसरों पर अध्यापकों को उन्नत शैक्षिक प्रविधियों की ओर ले जा सके। संकुल पर्यवेक्षकों का कर्तव्य है कि विद्यालयों में जाएं एवं अध्यापकों से शैक्षिक बिंदुओं पर चर्चा करें, अध्ययन-अध्यापन की प्रक्रिया को देखें एवं प्रभावकारी शैक्षिक विधियों को क्रियान्वित कराएं।

- प्रभावकारी विद्यालयी शिक्षा के लिए सतत अनुश्रवण (monitoring), सामग्रियों एवं प्रविधियों का पुनरीक्षण एवं प्रतिपुष्टि आवश्यक है। प्रभावशाली शिक्षण के लिए प्रेरणादायक प्रशिक्षण, मार्गदर्शन एवं अध्ययन-अध्यापन सामग्रियों का निर्माण आवश्यक है। संकुल शैक्षिक समन्वय मासिक गोष्ठियों में इन बिंदुओं पर ध्यान दें। ब्लाक, ज़िला एवं राज्य स्तरीय शैक्षिक प्रशासक शैक्षिक पर्यवेक्षक के रूप में अध्यापकों की शैक्षिक क्रियाओं में सहयोग करें, मार्गदर्शन करें एवं उन्नत विधियों एवं प्रविधियों को लागू करने में सहायता करें। इसके फलस्वरूप विद्यालय का शैक्षिक वातावरण, अध्यापकों की निष्ठा एवं सम्मान में उन्नति होगी।

इस प्रकार विद्यालयों की प्रभावशीलता बढ़ाने के लिए पर्यवेक्षक को विद्यालयों एवं अध्यापकों की कार्य परिस्थितियों को उन्नत करने में सहायता करनी होगी, पुस्तकों के निर्माण में सहयोग करके पाठ्यपुस्तकों के स्तर को उन्नत बनाना होगा, प्रभावशाली शिक्षण का उत्तरदायित्व लेना होगा एवं शिक्षकों को व्यावसायिक प्रोत्साहन देना होगा। पर्यवेक्षक शैक्षिक उपलब्धि में उन्नति निम्नलिखित प्रकार से कर सकता है :

- उत्साहवर्धक वातावरण प्रस्तुत करके,
- नियमित उपस्थिति को बढ़ावा देकर,
- विद्यालयों में गुणात्मक विकास करके,
- शिक्षण सामग्रियों के उपयोग के प्रति प्रेरित करके,
- अध्यापकों में गुणात्मक विकास करके,
- शिक्षण पद्धति को उन्नत करके अर्थात क्रियात्मक अधिगम, छात्रों के विकास का अनुश्रवण एवं प्रतिपुष्टि, अधिगम अभ्यास तथा छात्रों को कक्षा-क्रियाओं में भाग लेने के लिए प्रेरित करके।

साथ ही पर्यवेक्षक को देखना है कि छात्र क्या सीखते हैं, अध्यापक कितने निष्ठावान हैं, शिक्षा में माता-पिता एवं अभिभावकों की कितनी भागेदारी है एवं प्रधानाचार्य/प्रधानाध्यापक का नेतृत्व कैसा है। उन्हें शिक्षण-पर्यवेक्षण प्रदान करना है।

भारत में पर्यवेक्षण कार्य को प्रभावशाली बनाने के लिए एवं वास्तविक रूप में शैक्षिक प्रगति के लिए आवश्यक है कि पर्यवेक्षकों द्वारा 'फाइल वर्क' (संचिका कार्य) फार्म भरने के काम, पर्यवेक्षकों से अशैक्षिक कार्यों, तथ्य संकलन आदि के कार्यों को कम करना होगा। शैक्षणिक कार्यों, शिक्षण में प्रगति के मूल्यांकन एवं कक्षा गतिविधियों में गुणात्मक विकास हेतु प्रयत्नों पर बल देना होगा। विद्यालयों में कक्षा गतिविधियों के सुधार पर यथास्थल अध्यापकों की शिक्षण पद्धतियों एवं छात्र-अधिगम तथा उपलब्धि में उन्नति करने हेतु प्रशिक्षण देने के कार्य को महत्व देना होगा।

शैक्षिक प्रशासन या प्रबंध का मुख्य उद्देश्य शैक्षिक विकास ही है। अतः शिक्षा विभाग के सभी पदाधिकारी प्रशासनात्मक कार्यों में से कम से कम

सप्ताह में दो दिन अवश्य अपने कार्यालयों से बाहर निकल कर विद्यालयों का भ्रमण करें, शैक्षिक गतिविधियों का अवलोकन करें, कक्षाओं में छात्रों से क्रिया-प्रतिक्रिया करें, अपने अधीन शिक्षा अधिकारियों का निरीक्षण एवं पर्यवेक्षण इस प्रकार करें जिससे छात्रों की शैक्षिक उपलब्धि, विद्यालयों की प्रभावशीलता, पर्यवेक्षकों के शैक्षिक सुधार संबंधी प्रयासों का सही आकलन हो। पूरे शिक्षा-विभाग का ढांचा इसी हेतु निर्मित है कि अध्ययन-अध्यापन की प्रक्रिया में सुधार हो। अतः प्रशासनात्मक कार्यों एवं शैक्षणिक कार्यों में संतुलन लाना होगा। शिक्षा पदाधिकारियों की सभी गोष्ठियों में, विद्यालय निरीक्षण में एवं औपचारिक या अनौपचारिक चर्चाओं में कम से कम आधा समय शैक्षणिक विषयों पर चर्चा में लगाया जाए।

अध्यापकों के उन्मुखीकरण एवं प्रशिक्षण कार्यक्रमों में शैक्षिक पदाधिकारी भी उपस्थित हों एवं शिक्षण की जटिलता को सीखें एवं समझें। शैक्षिक प्रशासकों से अपेक्षा है कि वे शैक्षणिक प्रक्रियाओं, विधियों, अवधारणाओं, प्रवृत्तियों, अधिगम की जटिलताओं में इस प्रकार दक्ष हों कि इन क्षेत्रों में अध्यापकों को सहायता पहुंचा सकें, उनका मार्गदर्शन कर सकें। अतः शिक्षा के प्रत्येक नवाचार एवं नवीन प्रणालियों के उन्मुखीकरण एवं प्रशिक्षण का कार्य शिक्षा अधिकारियों एवं पर्यवेक्षकों से किया जाए जो अध्यापकों को उचित मार्गदर्शन दे सकें।

पर्यवेक्षकों के चयन में अच्छे अध्यापकों को प्राथमिकता दी जाए, विषयगत पर्यवेक्षकों की नियुक्ति की जाए, पर्यवेक्षकों का सतत प्रशिक्षण होता रहे तथा प्रशासनात्मक वातावरण की अपेक्षा शैक्षिक वातावरण निर्मित किया जाए। पर्यवेक्षकों को शैक्षिक नेतृत्व हेतु दक्ष बनाया जाए एवं उनकी उपलब्धियों के आधार पर प्रोन्नति की जाए। शिक्षा जगत से नौकरशाही समाप्त की जाए। शिक्षा के उच्चतम पदों पर नौकरशाह के स्थान पर योग्य एवं दक्ष शैक्षिक व्यक्तियों की नियुक्ति की जाए। प्रबंधात्मक कार्यों के लिए कुछ निम्न पदों पर प्रशिक्षित प्रशासकों की नियुक्ति की जा सकती है किंतु उनमें भी शैक्षिक सूझ एवं अनुभव अपेक्षित है।

पर्यवेक्षणिक नियंत्रण का संबंध क्या पढ़ाया जाए (पाठ्य पुस्तकें एवं पाठ्यक्रम), कब पढ़ाया जाए (छात्रों के विकास की अवस्थाएं), किसको पढ़ाया जाए (छात्र चयन), किस प्रकार पढ़ाया जाए (शिक्षण विधियां एवं गतिविधियां) तथा किस उद्देश्य से पढ़ाया जाए (छात्रों के सर्वांगीण विकास हेतु एवं सफल व्यक्तित्व के निर्माण हेतु) से है। अतः शैक्षिक क्रिया-प्रतिक्रिया को बढ़ावा देकर पर्यवेक्षकों को छात्रों के समुचित विकास में सहायक होना है। इस हेतु आवश्यक हो जाता है कि पर्यवेक्षण से प्राप्त अनुभवों के आधार पर पर्यवेक्षक अपनी कार्य प्रणाली में इस प्रकार परिवर्तन एवं विकास लाएं कि वह शैक्षिक विकास में सार्थक योगदान कर सकें। ❑

शिक्षा पर फीचर लेखन : आवश्यकता एवं महत्ता

बी. आर. परमार

कार्यक्रम अधिकारी
राष्ट्रीय अध्यापक शिक्षा परिषद्
दक्षिण क्षेत्रीय समिति, बंगलौर

सृजनात्मक लेखन में फीचर एक ऐसी आकर्षक विधा है जिसे आजकल लेखन के हर क्षेत्र में प्रयोग में लाया जा रहा है। शिक्षा में पत्रकारिता का अभी आरंभिक दौर ही चल रहा है। यदि शिक्षा क्षेत्र में फीचर विधा का व्यापक व प्रभावी प्रयोग किया जाए तो समस्याओं तथा उपलब्धियों को छात्रों, शिक्षकों, अभिभावकों तथा सरकार के समक्ष आकर्षक रूप में रखा जा सकता है। चिंतन स्तर को ऊंचा उठाकर नई सदी के लिए नई बहस शुरू की जा सकती है।

विदित है कि मानव एक जिज्ञासु प्राणी है और उसकी जिज्ञासु प्रवृत्ति का ही परिणाम है कि ज्ञान का संचयन, संवहन एवं वितरण प्रक्रिया का दिनों दिन विस्तार हो रहा है। साथ ही हो रहा है विज्ञान युग का आरंभ भी। मनुष्य प्रतिदिन अपने आस-पास के बारे में नई जानकारी चाहता है। इसके लिए वह तरह-तरह की पुस्तकें, पत्र-पत्रिकाएं पढ़ता है। रेडियो सुनता है। दूरदर्शन के कार्यक्रम तथा छवि-गृह में चलचित्र आदि देखता है। इन माध्यमों से प्राप्त जानकारी व्यक्ति के ज्ञान भंडार को बढ़ाती है। प्रत्येक प्रकार की लिखित अथवा मौखिक जानकारी शिक्षावर्धक ही होती है। लेकिन यहां शिक्षा पर फीचर लेखन का अभिप्राय विशिष्ट है। इसका मुख्य संबंध विद्यालयीन शिक्षा व्यवस्था से जुड़े विभिन्न पहलुओं से है, जिसमें औपचारिक, अनौपचारिक, प्रौढ़, सतत, जनजागरण तथा जनसंख्या शिक्षा आदि सभी शामिल किए जा सकते हैं।

प्रश्न उठता है कि यदि सभी प्रकार का ज्ञान शिक्षा का अंग है तथा किसी भी विषय पर लिखा गया लेख शैक्षिक लेख हो सकता है, तो फिर शिक्षा पर फीचर लेखन एक नया विषय या विधा कैसे हो सकता है। इस संदर्भ में यहां स्पष्टीकरण प्रासंगिक है कि शिक्षा फीचर लेखन नया विषय न होकर शिक्षा को नए दृष्टिकोण से सोचने-परखने का एक नया सृजनात्मक तरीका हो सकता है जिसमें शिक्षा जगत के उन पहलुओं को मार्मिक ढंग से वर्णित किया जा सकता है जिनको प्रायः लेखों में स्थान नहीं मिल पाता है अथवा लिखने वालों का उस ओर ध्यान गया ही नहीं है।

शिक्षा पर जो आलेख लिखे व छापे जाते हैं, वे

मुख्यतः या तो शोधपरक होते हैं या फिर शिक्षण विधियों पर टीका-टिप्पणी होते हैं। इनमें प्रायः एक ही शीर्षक के अंतर्गत कई आयामों का समावेश होता है जिससे लेखन और लेखक का उद्देश्य तो पूरा होता है, लेकिन कुछ आयाम या बिंदु शेष रह जाते हैं। फलतः उन आयामों व पहलुओं पर बहस नहीं छिड़ती जिनका शिक्षा सुधार से सीधा संबंध रहता है।

शिक्षा फीचर से अभिप्राय शिक्षा के विभिन्न पहलुओं पर लिखे जाने वाले फीचर से है। शिक्षा की व्यापक परिभाषा है - 'अंधकार से प्रकाश की ओर'। अतः बालक की व्यक्तिगत समस्या से लेकर अध्यापक के बुनियादी हक तक के सभी पहलू शिक्षा फीचर के विषय हो सकते हैं। बालक की अधिगम प्रक्रिया से संबंधित उसका व्यवहार, शिक्षक-पालक संपर्क, शिक्षक के कक्षागत अनुभव, प्राचार्य के व्यवहार का शिक्षण पर प्रभाव, अधिगम वातावरण निर्माण में विभिन्न घटकों की भूमिका, बालक का किसी एक विषय में अधिक रुझान अथवा अरुचि, ग्रामीण बालकों की अधिगम समस्या, रुचि अथवा अरुचि, हकलाने वाले, मूक-बधिर, विकलांग आदि बालक की शैक्षिक-सामाजिक समस्या, कालखंड मध्यांतर के समय छात्र वार्तालाप आदि ऐसे सैंकड़ों विषय हो सकते हैं जिन पर शिक्षा फीचर लिखे जा सकते हैं और जो शिक्षण प्रक्रिया को समझने में नया दृष्टिकोण तैयार कर सकते हैं।

आज आधुनिक संप्रेषण माध्यमों के कारण ज्ञान स्थापना के क्षेत्र में अध्यापक की भूमिका पर अनेक प्रश्न उठ रहे हैं। ज्ञान स्थापना के क्षेत्र में संचार माध्यमों की भूमिका दिन-प्रतिदिन लोकप्रिय होती जा रही है। लेकिन मानवीय संवेदना के अंकुरण-प्रस्फुटन का जहां तक सवाल है, यह जीवंत शिक्षण के माध्यम में ही संभव है। इन तथ्यों को ध्यान में रखते हुए लिखे जाने वाले फीचर को शिक्षा फीचर की संज्ञा देना अधिक सार्थक हो सकता है।

शिक्षा फीचर के प्रकार

शिक्षा पर लेखन आज के संदर्भ में बोझिल व नीरस पाठ्यक्रम, मूल्यांकन विधि के बारे में बदलती अवधारणा, नवाचार व नए-नए शोध का प्रभाव, शिक्षा केंद्र में शिक्षक के स्थान पर बालक, सतत मूल्यांकन प्रक्रिया का चलन, अध्यापक शिक्षा में विकास तथा न्यूनतम अधिगम स्तर के लिए अध्यापक को जवाबदेह बनाने जैसे विषयों के फलस्वरूप शिक्षा लेखन का महत्व बढ़ने लगा है तथा इन विषयों पर व्यापक चर्चा होने लगी है। वर्तमान में शिक्षक की आर्थिक, सामाजिक तथा राजनैतिक समस्याएं भी चर्चा का विषय बनी हुई हैं।

शिक्षा, शिक्षक तथा शिक्षण आदि की चर्चा को सार्थक रूप प्रदान करने के लिए उन तथ्यों को ईमानदारी से प्रस्तुत करने की चेष्टा की जानी चाहिए जिसमें पूर्वाग्रह नहीं होना चाहिए। विद्यालय कक्षा तथा अध्यापक का सही चित्रण ही नए आयाम को स्थापित करने में सहायक होता हैं। इन तथ्यों को ध्यान में रखकर शिक्षा फीचर के संभावित प्रकारों पर विचार किया जा सकता है।

विषय आधारित शिक्षा फीचर

- पाठ्यक्रम आधारित
- विद्यालय आधारित
- विद्यार्थी आधारित
- अध्यापक आधारित
- शिक्षण सामग्री आधारित
- शिक्षण विधि आधारित
- शिक्षण में नवाचार आधारित
- मूल्यांकन आधारित
- शिक्षा शास्त्र पर आधारित
- संतों-महापुरुषों की वाणी पर आधारित

वर्ग आधारित शिक्षा फीचर

- बाल वर्ग (कक्षा 1 से 5 तक के बालक)
- युवा वर्ग (कक्षा 6 से 12 तक के बालक)
- प्रौढ़ वर्ग (उच्च कक्षा तथा अनौपचारिक शिक्षा से संबंधित)
- विकलांग वर्ग (तीनों अवस्थाओं के छात्र-छात्राएं)
- बालिका वर्ग (तीनों अवस्थाओं की छात्राएं)
- शिक्षक वर्ग (प्राथमिक शिक्षा, माध्यमिक शिक्षा तथा उच्चतर माध्यमिक शिक्षा आदि में अध्यापनरत)
- अभिभावक वर्ग (शिक्षित अभिभावक, अशिक्षित अभिभावक, ग्रामीण क्षेत्र के अभिभावक, शहरी क्षेत्र के अभिभावक, उच्च आय वर्ग वाले अभिभावक, गरीब अभिभावक)

ध्येय आधारित शिक्षा लेखन

- व्यवसायोन्मुख शिक्षा फीचर
- सूचनापरक शिक्षा फीचर
- सकारात्मक शिक्षा फीचर
- नकारात्मक शिक्षा फीचर
- शिक्षा रीति-नीति शिक्षा फीचर

शिक्षा फीचर की विशेषताएं

शिक्षा आलेख प्रायः मासिक, त्रैमासिक अथवा अर्द्धवार्षिकी पत्रिकाओं में ही छपते हैं। दैनिक पत्र-पत्रिकाओं ने अभी शिक्षा लेखन को उल्लेखनीय महत्व नहीं दिया है। अतः पत्रिकाओं की सीमाओं को ध्यान में रखते हुए लेख इस प्रकार की विशेषता लिए होना चाहिए :

- गंभीरता
- सोद्देश्यता
- मौलिकता
- विश्वसनीयता
- सुसंबद्धता
- शिक्षा, शिक्षक व शिष्य के इर्द-गिर्द
- व्यक्तिगत आक्षेप का कोई स्थान नहीं
- सुधारात्मक चिंतन की प्रधानता
- तथ्यपरकता
- संयमित मत अभिव्यक्ति

शिक्षा फीचर की भाषा और शैली

अन्य लेखन की भांति शिक्षा फीचर लेखन में भी

भाषा व शैली का महत्वपूर्ण स्थान है क्योंकि लेख की सफलता भाषा व शैली पर भी निर्भर करती है। विषय की गंभीरता को ध्यान में रखते हुए प्रयुक्त भाषा व शैली पाठकों में साहस व विश्वास पैदा करती है। सरकार व शीर्ष संस्थाओं का ध्यान आकर्षित होता है और अंततः लेख में उठाए गए विषय पर गंभीरता से चर्चा होने की संभावना बढ़ती है।

शिक्षा लेखन के समय विशेष रूप से ध्यान देने योग्य बात यह होती है कि लेखन जिस स्तर अथवा समस्या पर होता है उसी स्तर अथवा समस्या के लिए प्रचलित शब्दावली का प्रयोग सरल, सुबोध तथा प्रभावोत्पादक रूप में होना चाहिए। शिक्षा पर कलम चलना एक संवेदनशील कार्य है तथा समाज व सरकार का इससे सीधा संबंध रहता है। अतः भ्रामक आंकड़ों तथा तथ्यों के समावेश से बचना चाहिए। विषय अधिक गंभीर, शिक्षण कार्य से सीधा जुड़ा हुआ होना चाहिए। साथ ही अच्छाई-बुराई के पक्ष को तटस्थता से रखने का प्रयास होना चाहिए। मुहावरों, लोकोक्तियों एवं कल्पना की ऊंची उड़ान तथा उच्च कोटि की साहित्यिकता का यथासंभव कम से कम प्रयोग किया जाना चाहिए।

शिक्षा पर फीचर लिखने की आवश्यकता

शिक्षा सौ तालों की एक चाबी है अर्थात जितने भी रास्ते अवरुद्ध हैं उनको खोलने के लिए शिक्षा रूपी चाबी का उपयोग किया जाता है और किया जाना चाहिए। संक्षेप में कहें कि इस संसार में जो भी प्रगति हुई है, हो रही है तथा होने की संभावना है, उन सभी के पीछे शिक्षा ही है। शिक्षा देश व जीवन पथ के लिए एक पहलू, क्षेत्र, दिशा, उद्देश्य, साध्य, साधन, आयाम तथा पहचान है। अतः शिक्षा पर फीचर लेखन की आवश्यकता स्वयं सिद्ध है।

लेकिन यहां शिक्षा फीचर की आवश्यकता भिन्न अर्थ में है। भारत सौ करोड़ की आबादी वाला देश है। यहां प्रतिदिन लगभग दस हजार पत्र-पत्रिकाएं प्रकाशित होती हैं। 700-800 दैनिक पत्र हैं। लगभग 500 श्रेष्ठ पत्रिकाएं मासिक अथवा त्रैमासिक रूप में छपती हैं। असंख्य लघु-पत्र-पत्रिकाएं छपती हैं जिनका प्रत्यक्ष संबंध किसी जाति, धर्म, भाषा, प्रांत, संप्रदाय, मत अथवा बोली से हो सकता है। संक्षेप में कहा जाए तो भारत में पत्र-पत्रिकाओं के प्रकाशन संबंधी प्रगति आशातीत है।

लेकिन जब प्रकाशित पत्र-पत्रिकाओं के कलेवर का अध्ययन करें तो ज्ञात होता है कि शिक्षा, शिक्षक, शिक्षण, विद्यालय, पाठ्यक्रम, प्रशिक्षण आदि पर लेखों अथवा समाचारों की संख्या का प्रतिशत लगभग नगण्य है। शिक्षा फीचर को पत्र-पत्रिकाओं में स्थान न देने अथवा न मिलने के अनेक कारण हो सकते हैं। लेकिन कुछ कारण स्पष्ट हैं जैसे– शिक्षा पर लेखन कार्य ईमानदारी से नहीं हो रहा है। शैक्षिक पत्रिकाएं प्रकाशित हो रही हैं लेकिन प्रकाशनार्थ सामग्री का अभाव है। आलेख लिखे जाते हैं लेकिन सच्चाई की कमी है, मौलिकता का अभाव है, सृजन क्षेत्र सीमित है। लेखक-पत्रकार अन्य विषयों पर लिखने में रुचि लेते हैं क्योंकि शिक्षा फीचर द्वारा प्रसिद्धि कम मिलती है। अतः यह निर्विवाद सत्य है कि शिक्षा जैसा क्षेत्र महत्वपूर्ण विषय होने के बावजूद दैनिक तथा मासिक पत्र-पत्रिकाओं में वह स्थान नहीं पा सका है जिसका

यह हकदार है। अतः शिक्षा जैसे उज्ज्वल, प्रासंगिक तथा उपादेय विषय के विभिन्न आयामों को प्रकाशित करने, प्रासंगिक बनाने तथा उपादेयता प्रदान करने के लिए शिक्षा पर फीचर लेखन की आवश्यकता है।

शिक्षा फीचर की महत्ता

ज्ञातव्य है कि शिक्षा सबसे महत्वपूर्ण विषय है। जब इस क्षेत्र में जुड़ी समस्याओं तथा उनके निवारण के तौर-तरीकों पर आलेख लेखन व प्रकाशन की संख्या बढ़ेगी और पाठक, अभिभावक, छात्र, शिक्षक, शिक्षक-प्रशिक्षक, योजनाकार व शिक्षा से संबंधित संस्थाएं एक दूसरे की समस्याओं तथा आवश्यकताओं से रू-ब-रू होंगे तभी सत्य का सही रूप सबके सामने आएगा। फलतः शिक्षा योजनाएं कागज़ों, फाइलों, व्याख्यान मालाओं, गोष्ठियों आदि तक सीमित न रह कर वास्तविक रूप में परिणित होंगी। बस्ते का बोझ कम होगा। मूल्यांकन विधियों में सुधार होगा। अभिभावक तथा समाज को विद्यालय को जानने का अवसर मिलेगा। जनता की शिक्षक के प्रति अवधारणा बदलेगी। अध्यापक की योग्यता में सुधार होगा। उच्च स्तर पर तत्परता से कार्यवाही होगी और इस तरह एक दिलचस्प शैक्षिक वातावरण निर्मित हो सकेगा जिसमें देश, समाज व मानव समुदाय के कल्याण की भावना निहित हो। ❑

पुस्तक समीक्षा

हिंदी शिक्षण—संकल्पना और प्रयोग

लेखक	: हीरालाल बाछोतिया
प्रकाशक	: आर्य प्रकाशन मंडल, दिल्ली
प्रकाशन वर्ष	: 1998
मूल्य	: रु. 54/-
समीक्षक	: हरीसिंह पाल विद्यालय प्रसारण आकाशवाणी, नई दिल्ली

मनुष्य एक प्रबुद्ध प्राणी माना जाता है। इसका कारण यह है कि वह सोच सकता है, समझ सकता है और अपने विचार व्यक्त कर सकता है। भाषा, संप्रेषण का एक सशक्त माध्यम है। सामाजिक व्यवहार और व्यावसायिक कौशलों के विकास के लिए भी भाषा अनिवार्य माध्यम है। भाषा व्यापक स्तर पर शासन और प्रशासन का भी सबलीकरण करती है। भाषा लिखित और शाब्दिक दोनों रूपों में प्रयुक्त होती है।

भारत में हिंदी भाषा को राष्ट्रभाषा के साथ-साथ राजभाषा का भी दर्जा प्राप्त है। अतः शासन, प्रशासन और शिक्षण-प्रशिक्षण कार्य में हिंदी भाषा का अप्रतिम स्थान है। यूं तो हिंदी भाषा पर अनेक पुस्तकें उपलब्ध हैं लेकिन अधिकांशतः ये पुस्तकें विभिन्न कक्षाओं के पाठ्यक्रमों पर आधारित हैं। कुछ पुस्तकें बी. एड., एम.एड. के पाठ्यक्रमों के लिए हैं तो कुछ प्रतियोगी परीक्षाओं के लिए। इस प्रकार हिंदी शिक्षण की समग्र प्रक्रिया पर स्तरीय पुस्तकों का एक प्रकार से अभाव-सा ही है। ऐसे में शिक्षाविद् डा. हीरालाल बाछोतिया की सद्यः प्रकाशित कृति 'हिंदी शिक्षण—संकल्पना और प्रयोग' इस अभाव को पूरा करती प्रतीत होती है।

हिंदी भाषा के शिक्षण की बारीकियों और उसके अनुप्रयोगों पर यह पुस्तक गंभीरता के साथ विवेचन प्रस्तुत करती है। इस कृति के प्रथम अध्याय 'भाषा' में भाषा के स्वरूप और उसकी संप्रेषणीयता पर प्रकाश डाला गया है। भारतीय विद्यालयों में हिंदी मातृभाषा तथा द्वितीय भाषा के रूप में पढ़ाई जा रही है। लेखक के अनुसार राष्ट्रभाषा हिंदी के माध्यम से ज्यों-ज्यों विभिन्न भाषा-भाषी लोग एक दूसरे को समझने का प्रयास करेंगे, वे एक-दूसरे की प्रादेशिक-सांस्कृतिक विशेषताओं को भी समझने लगेंगे। (पृष्ठ 14)

शिक्षण प्रविधियों के अंतर्गत इस कृति में संश्लेषण पद्धति, वर्णन विधि, व्यविता शिक्षण, संवाद एकांकी शिक्षण, व्याकरण शिक्षा और हिंदी शिक्षण के प्रकार्यात्मक पक्ष को सरल एवं सुगम विवेचन के साथ प्रस्तुत किया गया है। हिंदी के प्रकार्यात्मक स्वरूप को इन शैक्षिक भाषायी कार्यकलापों के स्वरूप के साथ समझाया गया है—भूमिका निर्वाह, नियंत्रित पुनर्कथन, वर्णन मुक्त कथन और भाषण, भाषिक खेल तथा लेखन कार्यक्रम। कुछ कार्यकलापों और उसके भाषायी रूपों की स्थिति, संवाद भाषिक खेल के माध्यम से रोचक उदाहरणों के साथ समझाई गई है।

पाठ्यक्रम के रूप में पढ़ाई जा रही हिंदी भाषा द्वितीय और हिंदी भाषा तृतीय की शिक्षण प्रविधियों और उनके प्रयोगों की व्याख्या इस कृति के तृतीय अध्याय में प्रस्तुत की गई है। हिंदी शिक्षण की प्रमुख विधियां हैं - व्याकरण, अनुवाद विधि, चयन विधि, वाचन विधि। आधारभूत वाक्य संरचना सिखाने के लिए स्थानापत्ति अभ्यास, अन्विति अभ्यास, रूपांतरण अभ्यास, प्रश्नोत्तर अभ्यास, वाक्य विस्तार अभ्यास, लिखित अभ्यास को सोदाहरण रखा गया है। पढ़ाने की विधि का भी विस्तृत वर्णन है।

अपनी सहजता, सरलता, सुबोधता और ग्राह्यता के कारण हिंदी एक लंबे अरसे से देश की लोकप्रिय भाषा रही है जो देश के अधिकांश क्षेत्र में समझी और बोली जाती है। हिंदीतर क्षेत्रों में भी हिंदी पढ़ने और पढ़ाने के प्रति पूर्ण ललक है। इस कृति के चतुर्थ अध्याय 'हिंदी शिक्षण : कुछ और मुद्दे' में लेखक ने राष्ट्र में उपलब्ध हिंदी शिक्षण सुविधाओं और उसके औचित्य पर गंभीरतापूर्ण विवेचन प्रस्तुत किया है। यह कटु सत्य भी पाठकों के समक्ष रखा है कि किसी भी स्तर या क्षेत्र में हिंदी शिक्षण की वर्तमान स्थिति पर संतोष व्यक्त करना, वस्तुस्थिति से आंख मूंदना होगा। केंद्रीय स्तर पर अध्यापन हेतु पठन पाठन सामग्री का निर्माण किया गया है। शिक्षा राज्य का विषय है अतः इस सामग्री का यथोचित उपयोग हो पा रहा है।

भाषा-शिक्षण के शैक्षिक संसाधनों के रूप में डॉ. बाछोतिया ने फ्लैश कार्ड, सही शब्द मिलान कार्ड, शब्द खेल, रिक्त स्थान पूर्ति, शब्द निर्माण को प्राथमिकता दी है। इलैक्ट्रॉनिक मीडिया के बारे में लेखक का मत है कि इसके माध्यम से शिक्षकों की शिक्षण विधि व प्रविधियों और पाठ्य पुस्तकों के विभिन्न घटकों की जानकारी दी जा सकती है।

लिपि वर्तनी मानकीकरण के पक्ष को लेखक ने बड़ी विद्वत्ता के साथ प्रस्तुत किया है। इसमें पहले लेखक और ध्वनि के संबंध का विवेचन है। इसके अंतर्गत खड़ी पाई के लोप, हलंत लगाने, संयुक्ताक्षर में 'र' के ती रूप अनुस्वार तथा अनुनासिकता के साथ-साथ विदेशी ध्वनियों की भी व्याख्या की गई है। लेखक ने माना है कि मुद्रण, टंकण और अब कंप्यूटर हमारी अभिव्यक्ति और भाषा को प्रभावित कर रहा है। इस दिशा में एकरूपता की पहल की धीमी गति को जितना तेज़ किया जा सके उतना ही अच्छा है।

'प्रयोजन मूलक हिंदी का स्वरूप' अध्याय में लेखक ने सामान्य हिंदी जो विभिन्न बोलियों से संपृक्त है, साहित्यिक हिंदी जो भाव पक्ष व कला पक्ष से जुड़ी है और व्यावसायिक हिंदी जो कामकाजी भाषा से संबंधित है, उनकी व्याख्या की है।

भारत में औद्योगीकरण के परिणामस्वरूप नगरीकरण का व्यापक क्षेत्र बना है। विभिन्न भाषा-भाषी लोग नौकरी के कारण एक साथ रहते हैं लेकिन शिक्षा माध्यम के रूप में उनके बच्चों को समस्या आती है। सरकार ने इसके लिए त्रिभाषा सूत्र के अंतर्गत हिंदी शिक्षण की व्यवस्था की है। त्रिभाषा सूत्र और हिंदी शिक्षण के तकनीकी पक्ष को लेखक ने अध्याय आठ में आवश्यक तथ्यों के साथ प्रस्तुत किया है। इस अध्याय में लेखक ने उपयोगी सिफारिशें भी की हैं जो व्यावहारिक भी हैं।

विद्यालय स्तर पर हिंदी शिक्षण तीन रूपों में मातृभाषा, भाषा दो और भाषा तीन के रूप में

किया जा रहा है। एक ही अध्यापक से हिंदी शिक्षण के तीन स्वरूपों का अध्यापन कराना न तो विषय के साथ न्याय है और न शिक्षार्थियों के हित में है। इसी तथ्य को लेखक ने 'हिंदी शिक्षण कुछ सैद्धांतिक सवाल' अध्याय में गंभीरतापूर्वक रेखांकित किया है। लेखक ने इस समस्या को शिक्षण तथा प्रशिक्षण, शिक्षण विधि और पाठ्य सामग्री के वर्गीकरण के आधार पर व्याख्यायित किया है।

इस कृति के परिशिष्ट में शिक्षा के अंतर्गत भाषाओं के स्थान पर व्यापक प्रकाश डाला गया है। लेखक का आकलन है कि—दुर्भाग्यवश 'त्रिभाषा सूत्र' को तोड़-मरोड़ कर पालन किया गया है। कुछ राज्य वास्तव में द्विभाषा सूत्र का अनुसरण कर रहे हैं। एक राज्य भाषायी अल्पसंख्यकों पर वास्तव में चार भाषा सूत्र थोप रहा है। हम इस मत के हैं कि देश की एकता और अखंडता के हित में सभी राज्यों में त्रिभाषा को समान रूप से सच्चे भाव से कार्यान्वित करने के लिए प्रभावी कदम उठाए जाने चाहिए।

द्वितीय भाषा के रूप में हिंदी शिक्षण पर बल देते हुए लेखक ने वातावरण निर्माण, सुनना और बोलना, पढ़ना-लिखना, सहपाठ्यक्रमी क्रियाएं आदि बिंदुओं पर विस्तृत प्रकाश डाला है। शिक्षक, भाषा शिक्षण के छात्रों को किस प्रकार प्रोत्साहित कर सकता है उसके लिए लेखक की मान्यता है कि छात्रों को उत्तर देना सिखाना चाहिए। साथ ही सही अभिव्यक्ति, पुनर्बलन, तादात्म्य स्थापन, सांस्कृतिक परिवेश, भाषा अधिगम पर बल देना आवश्यक है।

इस प्रकार यह हिंदी शिक्षण के सभी बिंदुओं को समग्रता के साथ प्रस्तुत करने में सफल है। जहां यह पुस्तक हिंदी अध्यापकों के लिए महत्वपूर्ण हैं वहीं यह हिंदी शिक्षण के प्रशासकों और नियोजकों के लिए भी समीचीन है। इस कृति में विषय को समझाने के लिए वांछित चित्रों, आंकड़ों और उदाहरणों का सहारा लिया गया है जो विषय की गंभीरता और शास्त्रीयता को सुगम बनाते हैं। हिंदी प्राध्यापकों, शिक्षकों और लेखकों को भी इस ग्रंथ के माध्यम से अपनी हिंदी सुधारने में काफी सहयोग मिलेगा।

राजभाषा विभाग को अपने शिक्षण प्रशिक्षण कार्यक्रमों में इस कृति से काफी संदर्भ सहयोग मिलेगा। आज जबकि हिंदी राजभाषा के रूप में पूरे देश में लोकप्रिय हो चुकी है ऐसे समय में हिंदी भाषा शिक्षकों के मार्गदर्शन के लिए इस पुस्तक की महत्ता और भी बढ़ जाती है।

कृति की भाषा में संप्रेषणीयता है, हिंदी भाषा के व्यावहारिक और मानकीकृत रूप को इसमें प्रस्तुत किया गया है। अहिंदी भाषी लोग जो सीख रहे हैं या हिंदी शिक्षण कार्य में लगे हुए हैं उनके लिए यह संदर्भ ग्रंथ के रूप में उपयोगी कृति है। अपवाद के रूप में एकाध शब्द इसकी सुगमता को भंग करता है जैसे— समाकलनात्मकता (पृ.101), यादृच्छिक (पृ.9), विकोडीकरण (पृ.10) आदि। इनके स्थान पर सरल और प्रचलित शब्दों का प्रयोग, अहिंदी भाषा क्षेत्र के पाठकों के लिए सुगम रहता।

कहा जा सकता है कि डॉ. बाछोतिया ने अपने विशद अनुभव और गंभीर अध्ययन के बल पर नई सोच के साथ एक अभिनव, उपयोगी और महत्वपूर्ण पुस्तक, हिंदी के पाठकों के लिए प्रस्तुत की है। ❑

3. *An ex post facto design,* in which some present effect is traced backward to an assumed causal complex of factors or forces at a prior date, using for this purpose such records as are available, since no new measures of the past can be made in the present, and relying as before on procedures of selective control. Studies of this sort appear in Chapter V.

Chapter III

CROSS-SECTIONAL EXPERIMENTAL DESIGN

ARE SOCIAL programs, or the treatment given one group of persons, more closely associated with acceptable personal adjustment than is found to be the case among a group of persons without such programs or treatment? A preliminary answer to this question, based upon the facts of a single date, may suggest the desirability of more extensive study over intervals of time Cross-sectional experimental design is a method of study of this problem which, if it reveals a relationship encouraging research of a more ambitious sort, serves the useful function of an exploratory effort.

The task of measuring relationships is performed by statistical methods when correlation coefficients are computed. But since the measure of association between two factors, with numerous other factors unknown, does not lead to reliable knowledge, we need to introduce control over these other factors in our observational procedures. This may be done by partial correlation analysis, which is a powerful tool for measuring the relationship between two factors while one or more other measured factors are held constant It is, however, not advisable to use partial correlation unless certain conditions of reliable initial measurements, homogeneity, sampling, and normality of distribution are met Hence resort is often made to some breakdown analysis[1] in which homogeneous sub-groups are sorted out and simple correlations then made within these sub-groups. When such an analysis yields high correlations we appear to have discovered a persistent or real relationship, which was obscured in the larger and more heterogeneous group.

Analysis by cross-sectional experimental design *makes homogeneous* the two groups to be compared—the experimental group

[1] Manual manipulation

which receives the program or treatment and the control group which is denied this program or treatment—and does so by matching the measurements on selected factors Sometimes this control is achieved by identical individual matching, but more often it is approximated by equating frequency distributions on a given trait. In either case, some degree of control of *known* factors is obtained so that any increase in the measured relationship of the two factors, treatment and adjustment, which results increases our knowledge of association but does not establish the existence of a cause-and-effect relationship. It may, however, provide additional evidence in support of a working hypothesis and thus encourage us to undertake the projected experimental design described in Chapter IV, as a procedure to discover if a cause-and-effect relationship exists. In any event, analysis of association by cross-sectional experimental design forces us to handle the data, and the matching process adds to our store of direct knowledge of the factors in the problem. This point has been argued with good effect by Ezekiel[2] and deserves more attention than it has received from research sociologists.

Two examples of cross-sectional analysis by experimental design may now be considered the personal adjustment of Boy Scouts in Minneapolis in 1938 and the personal adjustment of work relief clients compared with direct relief clients in St. Paul in 1939.

The Personal Adjustment of Boy Scouts[3]

Mandel[4] set out to analyze the relationship between the duration of Boy Scout tenure in the Minneapolis area and subsequent participation in community activities and in community adjustments of these Boy Scouts four years after leaving the organization. Large sums are spent annually in the belief that the program

[2] M Ezekiel, *Methods of Correlation Analysis* Wiley, 2nd ed , 1941; see also, for partial correlation, Margaret J Hagood, *Statistics for Sociologists*, Reynal & Hitchcock, 1941, pp 717–746, and G U Yule and M G. Kendall, *An Introduction to the Theory of Statistics*, Griffin, 1940, pp 261–287

[3] This section of the chapter was formerly published as part of an article, "Design for Social Experiments," in the *American Sociological Review*, December, 1938, pp 796–800

[4] "A Controlled Analysis of the Relationship of Boy Scout Tenure and Participation to Community Adjustment," M.A. thesis, University of Minnesota, July, 1938

of the Boy Scouts of America "builds character." The national officials directing the program are aware of this problem and have encouraged a number of studies to evaluate the results of the program.[5] The Mandel study differs from all other studies previously made in that it is a more thoroughgoing application of the experimental method Its results are consistent with earlier results in so far as comparability is possible. No attempt was made in the Mandel study to test the building of character. As we shall see, a simpler and more limited objective was studied.

The procedure was as follows From the 2050 cards in the 1934 drop-file of the Metropolitan Division of the Minneapolis Area Council, a random sample of every tenth case was taken This gave 205 individuals Next, an effort was made to follow up all of these former Scouts. A total of 102 were located and personally interviewed in 1938 The remaining 103 could not be located, had moved away, were deceased, or were not cooperative. The sample of 102 was then divided into groups. The first group had dropped out of scouting during the year 1934 with an average tenure of 1 3 years. This was called the control group of drop-out Scouts The second group consisted of Scouts who had completed an average of 4 years of tenure by the year 1934. This was called the experimental group of active Scouts Next, the two groups were paired by equating the frequency distributions on place of birth (rural or urban), father's occupation, health rating on the Bell test, and age-grade ratio (a rough rating on mental ability) Enforcing these four controls eliminated 22 individuals, so that the final control group of drop-out Scouts consisted of 40 boys, and the experimental group of active Scouts consisted also of 40 boys In the course of the interview in 1938, each of these 80 Scouts was scored for participation on the Social Participation Scale,[6] a scoring device developed by Chapin in 1927 and since then partially standardized for reliability and validity, on the Bell Adjustment Inventory,[7] and on the Rundquist-

[5] *Ibid*, pp 28–49, 86–89

[6] F Stuart Chapin, *Social Participation Scale*, University of Minnesota Press, 1937 (See Appendix B)

[7] High M Bell, *The Adjustment Inventory*, Stanford University Press, 1934

Sletto[8] scales of morale and general adjustment. These latter are all fully standardized scales with established reliabilities and validities Thus the Mandel experiment, although it used the crude control technique of equated frequency distributions rather than identical individual pairing, used as measures of adjustment scales that had been previously standardized and widely used.

The results of this experiment showed that the active Scouts of 1934 had a better adjustment in 1938 than the drop-out Scouts did on both the Bell and the Rundquist-Sletto adjustment scales. This characteristic is supplemented by a record of fewer delinquencies and delinquencies of less serious nature on the part of active Scouts than of drop-out Scouts Thus, the slight measured differences in adjustment (not statistically significant) are supported and reinforced by the evidence of non-quantified factors and demonstrate the value of the criterion of agreement among small differences.

There was, however, one marked difference of statistical significance, and that was the participation score of 11 00 points, the average for the 40 drop-out Scouts, and the participation score of 5 82 points, the average for the active Scouts. Here we find a difference of 5 18 between the two group means, or a drop-out group score which is 89 per cent higher than the active group score in 1938 Mandel points out that this difference, which is statistically significant, may be due to the desire of individual boys in the drop-out group not to "channelize" their participation into scouting alone, but to diversify their participation in other activities Active Scouts seem to concentrate all of their interests in one activity The higher social participation score of the drop-out Scouts may be explained as due to (1) their search for a basis of integration for their individual desires, since the active Scouts may have found already *within* the varied program of the Scout organization a framework of integration not easily possible from participation in diverse community groups; or (2) their wish not *to narrow down* their activities to one type of organization at such an early age, motivated by the *wish* for new experience and

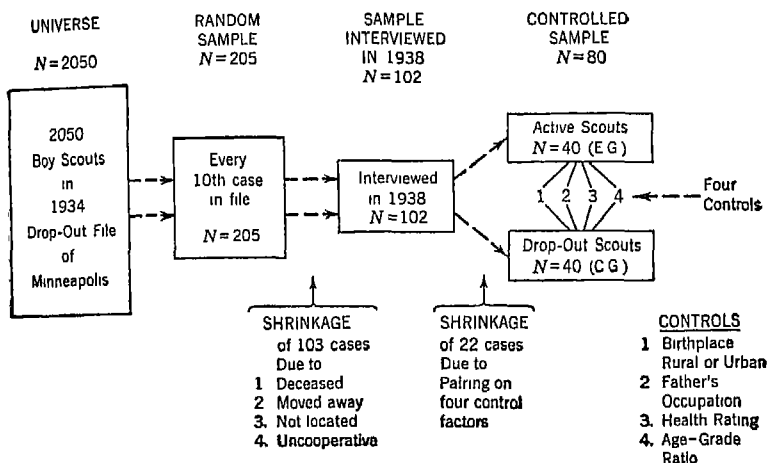
[8] E A. Rundquist and R F Sletto, *Minnesota Survey of Opinions* (short form), 1936, pp 394–398 in *Personality in the Depression,* University of Minnesota Press, 1936.

the *wish* for response, which they satisfied in taking part in a variety of activities; or (3) their *greater intensity* (more memberships on committees and more official positions held) as well as *greater extensity* (memberships in a larger number of organizations); or (4) the fact that they were more *socially acceptable* than active Scouts to admission into a large variety of community activities. Since the drop-outs were a more variable group and more non-conformist in behavior, it seems probable that explanations (1), (2), and (3) are true, rather than (4). The fact that the drop-outs were slightly less well adjusted on the Bell and Rundquist-Sletto scale norms tends also to minimize the value of explanation (4). In general, all the evidence of this experiment tends to show that boys who had stayed four years in scouting and thus secured the full benefit of the scout program up to the year 1934 were on the whole a more conventional, conforming, and homogeneous group in 1938 than the drop-out group The latter was more variable, less conventional, and more non-conformist in 1938.

If it be accepted that the valid measure of adjustment is based upon attitudinal factors expressed in verbal responses measured by the Bell and Rundquist-Sletto scales, then the criterion of agreement among small differences is in favor of the active Scouts On the other hand, if it be held that evidence of overt participation in community activities is a better measure of community adjustment, then the strikingly higher score of the drop-out Scouts indicates that they are the better adjusted The final conclusion is then a matter of judgment. Evidently it is necessary to do additional research on this problem. It would be advantageous to repeat the experiment using similar techniques under like conditions.

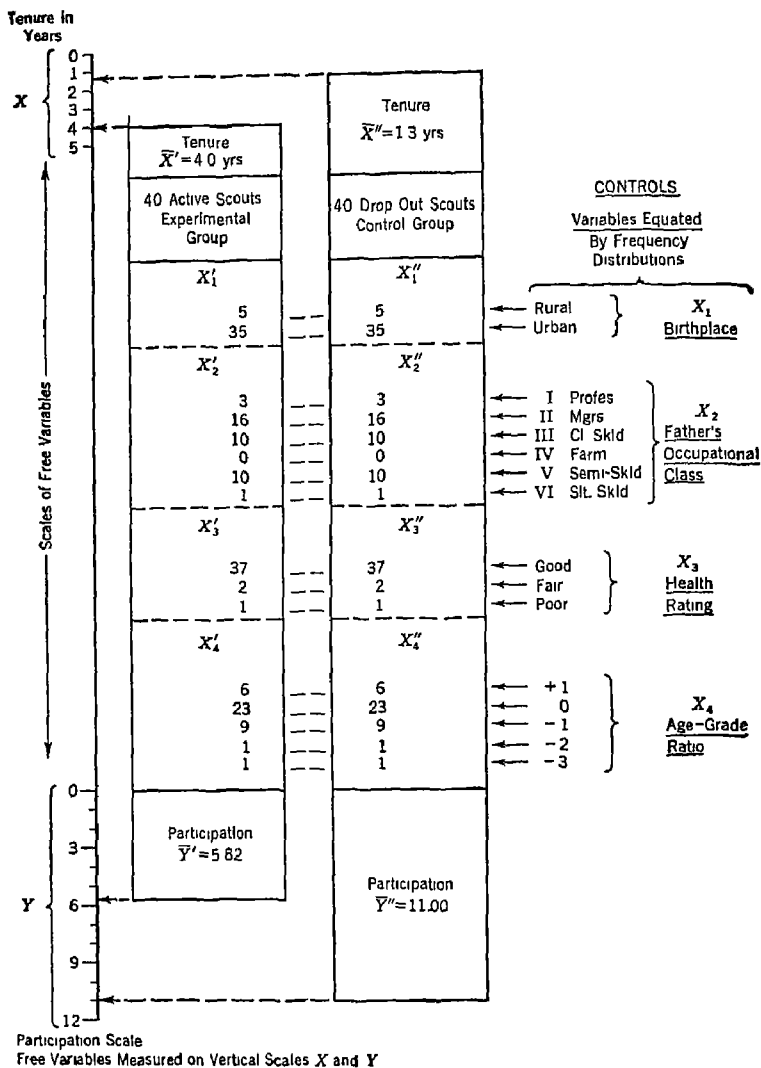
Our design for social experiments may be summarized in visual form. We present this summary in Figures 1 and 2 using some of the data of the Mandel experiment to make the exhibit definitive. Figure 1 shows the general procedure followed in the use of the experimental method and indicates graphically the successive steps and the corresponding shrinkage in the size of sample observed. Figure 2 shows the more detailed technical aspects of using the

technique of control by equating the frequency distributions of four variables which it is desired to hold constant. It also shows the relationship between the measured free variables Thus, longer Scout tenure (an average of 4 years for the experimental group) is associated with a significantly lower score on social participation,

FIGURE 1. Design for Social Experiments[a]

[a] Based on Mandel data.

and shorter Scout tenure (an average of 1.3 years for the control group of drop-outs) is associated with a social participation score that is 89 per cent higher The slight differences in the adjustment scores are not shown on the chart because of the difficulty of representing four-dimensional variation on a flat surface. Nevertheless, the charts serve to present the main outlines of a design for social experiment. It may be remarked parenthetically that this design for experiment is substantially the same as the schematic pattern we presented in 1931 and subsequently published in *Social Forces*.[9]

[9] F Stuart Chapin, "The Advantages of Experimental Sociology in the Study of Family Group Patterns," *Social Forces*, December, 1932

FIGURE 2. Design for Social Experiments[a]

[a] Based on Mandel data.

The Personal Adjustment of Work Relief Clients [10]

A program of work projects for the able-bodied, needy unemployed is generally assumed to have two important advantages over a system of direct relief. In the first place, it is obvious that such a program results in additions to the facilities and services of a community—public buildings, roads, educational activities, etc.—which would not be possible otherwise. The second advantage involves the maintenance and development of skills, work habits, and morale among the unemployed.

Despite the fact that the value of a work relief program in maintaining the morale of the unemployed is widely accepted, objective evidence on the question is fragmentary. This study attempts to provide an answer to the question, Does work relief maintain a higher morale than direct relief? Since an investigation of this nature concerns a study of psychological factors, and since these factors must be spontaneously expressed if they are to reveal actual differences in morale, this investigation was carried on quietly and without publicity.

The morale of eighty WPA workers in St. Paul was found to be 5.46 per cent higher than the morale of a matched group of 42 direct-relief clients in St. Paul (certified for WPA and waiting assignment). Besides measurement of morale, we obtained a pattern of adjustment by using three additional criteria—general adjustment, social participation, and social status. The average percentage superiority of the WPA workers over the matched group of direct-relief clients was 5.67 per cent on the percentage gains of these four criteria. Assuming these four criteria of adjustment to be a pattern that is more reliable than any one of the factors considered alone, we may ask: Is the difference in pattern of adjustment one that may occur so often by chance that it is merely a sampling error and not significant?

[10] This portion of the chapter was originally published in an article by the author and J. A. Jahn, "The Advantages of Work Relief over Direct Relief in Maintaining Morale in St. Paul in 1939," in the *American Journal of Sociology*, July, 1940, pp. 13–22. The Division of Research of the Work Projects Administration, which cooperated in making records available and in supplying information on administrative procedures, has given us permission to release the results of this study.

We cannot, of course, claim that the superiority of the WPA workers studied was exclusively the effect of the WPA program as a cause. But we can answer the question: What is the probability of obtaining

TABLE 3 Differences on Four Criteria of Adjustment WPA and Direct-Relief Groups

Groups Compared	Mean Scores on Four Criteria of Adjustment			
	Morale[a]	General Adjustment[a]	Social Participation	Social Status
WPA group, 80 cases	64.01	48.41	4.94	66.29
Direct-relief group, 42 cases	67.71	50.98	4.67	62.29
Differences between mean scores	− 3.70	− 2.57	0.27	4.00
Percentage differences	5.46	5.04	5.78	6.42
Critical ratios of these differences[b]	− 1.92	− 1.73	0.20	0.81
Combined critical ratios of four differences[c]		2.33	Odds against chance occurrence, 49:1	

[a] Scores on morale and general adjustment are on reverse scales, i.e., the higher the score the worse the morale, hence a negative difference is a gain. On the last two scales an increase in score is a gain.

[b] Computed from Fisher's formula.

[c] Computed from Guttman's formula, "On Uses of the Critical Ratio," by Louis Guttman, a research paper for the M.A. degree, University of Minnesota, June, 1939, pp. 36–40. Guttman proves that although the means of two variables are correlated in random samples, nevertheless, in random samples differences between pairs of means of one variable correlate zero with differences between pairs of means of the other variable. Thus the formula for the multiple critical ratio (M.C.R.) is

$$M.C.R. = \frac{c_1 + c_2 + \cdots + c_m}{\sqrt{m}}$$

Since there are several items common to the general adjustment scale and the morale scale, the combined critical ratio of 2.33 is excessive and somewhat spurious. When allowance is made for these factors, the combined critical ratio reduces to 1.7, or odds of about 9 to 1.

a difference, of the magnitude found, by pure chance? The odds are 9 to 1 of obtaining by chance a difference in pattern of adjustment as large as the advantage that the 80 WPA workers showed over a matched group of 42 direct-relief clients. Clearly, such a difference in pattern of social adjustment does not often occur by chance; there-

fore, we may conclude that the higher morale, general adjustment, social participation, and social status are perhaps due to the WPA program. In other words, we have some real evidence to support the conviction that a work program is better than direct relief in maintaining morale

We must next try to answer such questions as the following: (1) Were the WPA group and the relief group studied selected groups that would show differences in morale anyway? (2) How can you know that the differences found were not due to other factors, since adjustment is at best a very complicated process? (3) What dependence can you put upon the measurements of morale; would not some other investigator have obtained a different result? (4) Can you rely for a description of differences in complex adjustment upon only four criteria, such as those used?

Consider the first question· (1) Were the WPA group and the relief group studied selected groups that would show differences in morale anyway? In March, 1939, when the study was begun, there were 8074 at work on WPA in St. Paul on the twenty-ninth of the month. At the same time there were only 465 cases in all on direct relief certified as employable and waiting assignment to WPA Obviously, the number we had to choose from in the relief list was comparatively small. Our first step was to select a 5-percent sample at random from the WPA lists, to be compared with the direct-relief group In order to simplify the problem of comparison, there were eliminated from the two samples all persons previously on WPA but now on the waiting list, all single persons, and all divorced, separated, and widowed cases Also all persons not living in St. Paul and not working in Ramsey County were excluded. The result of equating the two groups, the WPA group and the direct-relief group, on these seven factors was to make all the individuals selected for study more comparable than would otherwise have been the case

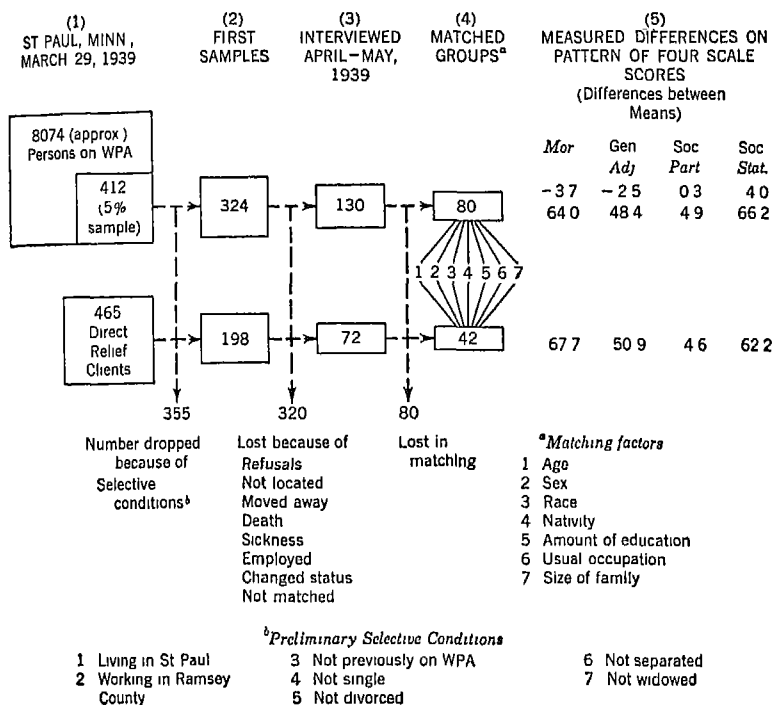
The next step was to eliminate from the two groups those individuals who still showed marked differences on other factors in addition to the seven mentioned above. Since marked differences in the distribution by former occupation, size of family, age, sex, years

of education, race, and nativity might affect morale, the two groups were matched in respect to each of these seven additional factors. In this way, extreme variations were eliminated, and the two groups were made as similar as possible in all characteristics except morale and adjustment, which might be associated with being on WPA in one group and with receiving direct relief in the other. Thus our method of dealing with question (1) was to apply the familiar control-group procedure and so protect our results from the limitations that would arise from "selected" samples of "bias" in the groups to be compared. This process did, however, reduce the number of cases for interview, so that at the time the field work was begun, we had only 172 WPA and 102 relief cases, of which 130 WPA and 72 relief cases were interviewed.

Social workers and graduate students in social work then visited these persons in their homes and obtained from each person responses to the Rundquist-Sletto *Survey of Opinions* (from which the morale score and the general adjustment score of each individual was later computed) and also information on the *Social Participation Scale* and the *Social Status Scale.* In this process of interviewing additional cases were eliminated because of refusals, not located, moved out of town, changes in relief status, separated from relief, sickness and death, or working in private employment. The net result was the final groups, 80 WPA workers and 42 direct-relief clients, complete as to information on each of the four criteria used to measure adjustment and matched on 14 characteristic traits that were the ascertainable facts of their social situations (see Figure 3).

Turning now to the second question, (2) How can you know that the differences found were not due to other factors, since adjustment is at best a very complicated process? we can answer by stating that since we matched on seven initial factors (dwelling residence, work residence, previous WPA experience, single, divorced, separated, or widowed) and again on seven additional controlled factors (age, sex, size of family, former occupation, years of education, race, and nativity), the differences found in favor of WPA workers as compared with direct-relief clients could not have been due to these fourteen factors Now it is apparent, of course, that

factors upon which information was not obtainable could not be held constant by matching. Such factors will, no doubt, occur to the critical reader All that we can say in extenuation is that we were limited by the available information and that another experiment

FIGURE 3. Flow Chart of Process of Study

might be improved by more complete coverage One such factor is income. This factor would have to be controlled if we desired a more precise isolation of the single factor, work, to determine how it alone is related to adjustment. Disparity of income as between the WPA and relief groups, otherwise matched on 14 factors, is not as serious a limitation as might first appear In a previous study, we found that income tended to vary with the class of occupation and

with the number of years of education Consequently, when variation in occupation and variation in education were controlled, the variation in income was strikingly diminished. Since in the present WPA study we did control occupation and education, there is good reason to believe that income also was controlled to a marked extent. This evidence is indirect but is probably quite significant Finally, it may be claimed with some point that larger income is a necessary part of the WPA program.

Another factor not controlled in obtaining the first results was length of time on relief.[11] By further elimination of cases in matched comparisons, we were able, however, to control this eighth factor.

TABLE 4 Comparisons of the Morale, General Adjustment, Social Participation, and Social Status Between WPA Workers and Relief Clients, Holding Constant Sex, Race, Nativity, Usual Occupation, Age, Education, Size of Family, and Length of Time on Relief

Groups Compared, Differences and Constants	Variables Compared			
	Morale [a]	General Adjustment [a]	Social Participation	Social Status
Means				
1 WPA group (n = 37)	65 27	49 54	5 19	62 78
2 Relief group (n = 25)	66 76	50 44	5 84	59 44
Differences between means	−1 49	−0 90	−0 65	+3 34
Critical ratios (Fisher's formula)	−0 56	−0 46	−0 32	+0 47

[a] Lower morale and adjustment scores on the Rundquist-Sletto test indicate higher morale and adjustment in the customary meaning of these terms

This process reduced the final numbers to 37 WPA and 25 direct-relief cases In this final comparison the WPA group still showed a better pattern of adjustment on the whole than the relief group, although the absolute differences on each of the four measures of

[11] Time on relief means here the length of time since the subject's family was registered for the first time with the direct-relief agency All WPA workers in St Paul were so registered before being placed on WPA projects

adjustment were smaller than in the case of comparisons using seven controlled factors[12] (see Table 4).

The next question now deserves attention (3) What dependence can you put upon the measurements of morale; would not some other investigator have obtained a different result? The answer to this question is the simplest of the list It is this Morale was measured by responses to a standardized scale constructed to measure morale. By a standardized scale we mean a measuring instrument that has been tested hundreds of times and does, in fact, give almost identical results when used by different investigators on the same or similar groups of subjects. Thus, the morale scale is a reliable scale. It is also a valid scale It does measure what it is designed to measure because it differentiates employed and unemployed persons and other contrasting groups which are ordinarily taken as typical of differences in morale by common-sense judgments of competent persons. The same conditions of reliability and validity are met by the other three scales general adjustment, social participation, and social status

Since the critical reader may wish to know additional facts about the method of using these four scales in interviews, the following description may anticipate further questions.

The scales measure (1) *morale,* or the degree that the individual feels competent to cope with the future and to achieve his desired goals, (2) *general adjustment,* or the feelings about his relationship to other persons, toward present or future social conditions, and toward present social institutions; (3) *social participation,* or the degree to which an individual actually engages in the organized activities of his community in terms of membership, attendance, contributions, committees, and offices, and (4) *social status,* or the position the family occupies with reference to the average prevailing household possessions of other families in the community.

It will be noted that the scales to measure morale and general adjustment describe *how people feel* about their situation, whereas the scales that measure social participation and social status de-

[12] These decreases in differences probably reflected fluctuations, owing to the small number of cases in these samples remaining after adding time on relief as a control

scribe their *objective social condition*. Together, the four scales enable us to describe the pattern of their adjustment to new or to old surroundings Since each scale has been in use and tested out hundreds of times in a number of social investigations, there are published norms available for comparison For example, morale scores of 58 to 59 points obtained from slum populations in another 1939 study are decidedly better than 65 for WPA and 67 for relief clients studied at the same time (see Table 3) and not far from those of the normal population. The general adjustment scores of 43 points in the same slum study are similarly better than 49 for the WPA and 50 for relief clients, and again not far from those of the normal population. The participation scores of the WPA and relief clients are worse than those of slum families in a previous study of the same area. Social status scores of 57 and 59 points in the same slum study compare unfavorably with those of the WPA (63) and of the relief group (59). Previous studies show, however, that a score in the fifties is within the observed range of normal working-class families in Minneapolis, Massachusetts, and Alabama.

The material gathered in the interview consisted of:

Minnesota Survey of Opinions, two sheets with 31 questions about the individual's attitudes, to be filled in by the subject After the interview, the *moral score* and the *general adjustment score* may be extracted from the subject's marked response by a simple system of weighing and scoring It takes the subject from 20 to 30 minutes to fill this in [13]

Social Participation Scale, one sheet for entries on each group affiliation of subject recorded in five entries under five columns by the visitor in reply to questions answered by the subject It takes 10 or 15 minutes to fill in the subject's answers [14]

Social Status Scale, one sheet containing 21 entries filled in as observations made by the visitor, with perhaps one or two non-inquisitorial questions. Can be completed in 5 minutes' observation

[13] These scales and their norms will be found in E A. Rundquist and R. F Sletto, *Personality in the Depression*, University of Minnesota Press, 1936

[14] These scales and their norms will be found in F S Chapin, *Contemporary American Institutions*, Harper & Brothers, 1935, and 1946 ed , reprinted by William C. Brown, Dubuque, Ia , pp 373–397, and F S Chapin, "Social Participation and Social Intelligence," *American Sociological Review*, April, 1939, pp 157–166.

Finally, the reader may ask question (4), Can you rely for a description of differences in complex adjustment upon only four criteria, such as those used? This is a difficult question and only a qualified answer can be given, an answer which the reader may accept or not, according to his own judgment. The answer can be made only in terms of several considerations In the first place, the four scales have differentiated between Negroes and Jews in a slum population study in Minneapolis in 1935–1936, between overcrowded slum families and not overcrowded slum families,[15] and between former Boy Scouts of four-year tenure and Boy Scouts with two-year tenure in 1938.[16] In the second place, the probability of a chance difference in pattern of adjustment between WPA workers and direct-relief clients is 9 against 1, whereas the highest chance occurrence for any single criterion (morale) is one in which the odds are only 17 to 1. Obviously, a pattern of differences on four factors (morale, general adjustment, social participation, and social status) is less likely to occur by chance in random sampling than is a difference on any single factor Hence when such a difference in pattern does occur, we may conclude that it probably represents a real difference between the WPA and direct-relief groups studied [17]

The chief purpose of this study was to discover whether a better

[15] F S Chapin, "The Effects of Slum Clearance on Family and Community Relationships in Minneapolis in 1935–1936," *American Journal of Sociology*, March, 1938, pp 744–763

[16] N G Mandel, "A Controlled Analysis of the Relationship of Boy Scout Tenure and Participation in Community Adjustment," M A thesis, University of Minnesota, July, 1938, and preceding section in this chapter

[17] The differences in favor of WPA workers found in this study were obtained several weeks before the strikes of July, 1939, and so could not have been influenced by these disturbances Utmost care was taken to allay suspicion by making it clear at the beginning of each interview that we were asking cooperation in obtaining people's opinions as part of a larger study conducted by a sociologist at the University of Minnesota It was made clear that this study had no bearing on WPA operations. Although the results of this study were favorable to the WPA program in St Paul as of the spring of 1939, every precaution was taken to secure objectivity and to avoid bias Although the WPA Division of Research expressed an interest in the results, the collection and interpretation of the facts were carried out entirely by us without suggestions or restrictions, except concerning administrative matters related to the use of WPA official records Under these circumstances the discovery of results favorable to WPA is all the more significant.

adjustment of WPA workers than a comparable group of direct-relief clients could be established as caused by the Works program. The outcome of this single and somewhat limited study does establish the presumption of such a cause-and-effect relationship Social workers are justly skeptical of simple cause-and-effect explanations in the complex social life of today It is wisely observed that a large number of influences combine to explain each specific social situation And yet, by matched groupings, we have succeeded in securing some control over fourteen factors related to the etiology of individual adjustment. When these fourteen factors are held constant, we discover that in St. Paul the Works program is associated with higher adjustment and direct relief is associated with lower adjustment. We cannot arbitrarily say whether the absolute differences found on measures of adjustment are important. We can and do state that there is a slight probability against the chance occurrence of absolute differences of the magnitudes actually found When we say "chance," we mean a large number of small causes; for "chance" is a term which never means, in the language of science, the absence of cause. When the odds are heavy against a difference being due to chance (i e., due to many small influences or causes), then we may claim that the differences are probably due to a few very influential factors, or causes, in this case the WPA program as contrasted to the direct-relief program.

How far these favorable results from the 1939 St. Paul study may be generalized to WPA at large, it is not possible to state Obviously, the next step is to extend this sort of study to other areas, repeat the procedures, and discover if the results are verified or disproved.

In passing from this chapter to a description of projected experimental design in Chapter IV, and to ex post facto experimental design in Chapter V, we make a transition from analysis of association between two factors at one date to the method of controlled observation of changes in this association that occur over an interval of time. Here we make the attempt to discover the type of association that is called cause and effect, or concomitant variation.

Chapter IV

PROJECTED EXPERIMENTAL DESIGN. THE CLASSICAL PATTERN OF "BEFORE" AND "AFTER" EXPERIMENTS THAT OPERATE FROM THE PRESENT TO THE FUTURE

AMELIORATION of the problems of human relations often takes the form of social programs of treatment to achieve desired ends But interwoven with the means or the social programs to attain chosen objectives there are other individual and social factors which combine to effect the results The method of projected experimental design provides a technique of study which we believe will enable us, as the technique is improved, to separate the planned means-ends schema from the more impersonal cause-effect relationships, so that we may assign some measure of relative influence to each It is the purpose of this chapter to analyze four studies the effects of a program of rural hygiene in Syria, the social effects of public housing in Minneapolis; the results of a student counseling program at the University of Wisconsin, and the consequences of a program of social treatment on juvenile delinquents in New York City. Before examining these separate studies, a brief consideration of the concept of cause and effect seems indicated.[1]

[1] An extensive treatment of the concept will be found in R M MacIver's *Social Causation*, Ginn, 1942, and R B Lindsay's article, "Causality in the Physical World," *Scientific Monthly*, October, 1933, pp 330–337, is an excellent brief summary The concept of prediction is closely related to the concept of causation The former was discussed in chap 1 of this book, especially in its semantic connotations For an extended discussion see Paul Horst *et al*, *The Prediction of Personal Adjustment*, Bulletin 48, Social Science Research Council, 1941 It is desirable to distinguish between the concept of estimation and the concept of prediction The former relates to the calculation of values of a dependent variable by substitution in a regression equation of the successive values of the independent variable or variables and does not necessarily include the time factor, which is a fundamental element in the concept of prediction.

Cause and effect, or causality as a system of ideas, is an explanation of successive events by a set of assumed antecedent-consequent relationships for which the evidence is objective and recurrent. In the area of physical things this type of explanation is quite generally accepted as a true explanation of what happens. In fact, predictions can be made with considerable confidence in narrow areas of experience, and in mechanics, where control is quite complete, determinism reigns But in the area of human relations explanations of social events in terms of cause and effect have made only crude beginnings, and more often than not are rejected as entirely inapplicable. But let us consider the alternatives.

The alternatives to acceptance of the causality explanation of human relations are to accept: (1) chaos, (2) magic, demonism, and supernaturalism, or (3) means-ends explanations Since chaos is no explanation at all, but a confession of the permanence of ignorance and of the persistence of impotence, it has to be discarded in any effort at rational explanation of social relations Similarly, magic, demonism, metaphysics, and supernaturalism have failed to explain the problems of social relations in a rational manner, albeit such systems of thought still, unfortunately, influence human relations. Even the means-ends explanation is only a partial answer to our questions, since many social events, like inflation, a run on a bank, and world war, defy explanation in the simple terms of purpose, design, and planning by any single or preponderate means-ends schema Nevertheless it is perfectly evident that narrowly conceived means-ends schemata do produce social effects, as when labor unions resort to strike to better the working conditions of their members, or when huge industries plan to achieve monopoly. But the situation is not so simple in other cases. When different interest groups in a community pursue their own independent means-ends plans without regard for consequences to the larger public, there are unplanned results that come from the combinations of these independently planned social actions. Such seems to be the case in the problem of inflation, and for the global community this is the case in the problem of world wars. Although the causal explanation of the

problems of inflation and of world wars is now based upon insufficient knowledge of the many factors at work, it does have the advantage of a courageous effort to seek understanding which strives on the one hand to avoid escape from reality into phantasy and chaos, and on the other hand to avoid yielding to the blandishments of magic, demonism, and supernaturalism The concept of causality when frankly empirical helps to order our thinking about complex problems to a logical system susceptible of test by facts.

The Concept of Causality: Cause and Effect

The concept of cause and effect as used in this book, therefore, does not mean rigid determinism Rather it is used as a shorthand device to represent a kind of association between factors in time sequence which has a determinable probability of occurrence Cause and effect is useful terminology since it is a more objectifiable mode of verbal expression than the folk language concept of means-ends. In the case of means and ends we are concerned primarily with preferred means to attain objectives that are desired.

By substituting the concept of cause and effect we avoid undue emphasis on the function of language to *express feeling* and emotion by turning attention to its *representative function.* This is achieved by resort to observation of how things happen Is the factor X (treatment) associated with the factor Y (adjustment)? This is a question that can be answered by correlation analysis once we have developed reliable scales to measure degrees of treatment and degrees of adjustment, that is, quantitative description of X and Y. The next step is to find out whether X precedes Y by an interval of time that can be described in the operations of measurement. When we find that changes in the magnitude of Y (measured adjustment), between a date when X (treatment) was begun and a second date when treatment was terminated, are changes in magnitude which occur very infrequently in chance, or in the fluctuations of random samples (the control group) selected from a non-treated universe of persons in other respects similar to those receiving treatment, we conclude that a calculable degree of probability supports

the hypothesis that extended treatment causes changes in personal adjustment.

Thus defined, in terms of operations performed, to describe how events happen, an element of objectivity is introduced into our procedure, i e, two or more equally competent observers may check one another's records because the terms used in describing what happens represent differences in kind or preferably of degree (measurement on a scale) of attributes (treatment and adjustment) of the subjects studied, and the terms do not express merely how the observer *feels* about that which he observes We fall into the latter situation when our account of treatment states that an appeal to the better nature of a delinquent boy can and does lead to an improvement in his conduct By using such words as "better" and "improvement" we disclose the fact that the observer has an emotional set or belief or preference for certain forms of behavior or states of being as against others. The objection to the latter type of verbal expression is the difficulty of obtaining agreement on the meaning of "better" or "improvement" *after* these terms have been used When we substitute for these normative terms the concept of change in mean score on a reliable scale that measures adjustment,[2] we have a statement that can be verified, since equally competent observers using the same scale on the same subjects under the same conditions would obtain results differing only in terms of small errors of observation, the frequency of which would tend to occur in approximation to some probability distribution Thus the purpose of "before" and "after" measurements is to supply an operation which adds to the objectivity of description, it is not a commitment to an assumption of rigid determinism.

[2] This is the question of validity of the scale, and we assume that the scale in question discriminates between individuals in degrees that are consistent with some independent criterion of the same attribute But even *before* validation of the completed scale as above, an element of validity is introduced when the component items of the proposed scale are rated independently by judges whose opinions are generally accepted in the community as discriminating judgments, thus providing an initial agreement on the meaning of items, a procedure in sharp contrast to that in which statements of interested practitioners who are working on the case are taken at their face value

The Effects of Rural Hygiene in Syria

One of the best documented and recorded results of an experiment that has been published is Stuart C. Dodd's, *A Controlled Experiment on Rural Hygiene in Syria.*[3] Dodd set out to discover the relationship between a program of rural hygiene and the hygienic practices of the families that were supposed to benefit by the program. The program of education in hygiene was put on by an itinerant traveling clinic of the Near East Foundation from 1931 to 1933. The beneficiaries were 40 families in the Arab village of Jib Ramli in Syria It is generally believed that efforts to improve the hygienic practices of individuals and families actually result in such improvement that morbidity and mortality are lowered, and general comfort and happiness is increased. Strange as it may seem, the validity of this assumption has seldom been tested, although millions of dollars are spent annually throughout the world on preventive medicine.

Dodd decided to set up an experiment that would attempt to measure the effects of hygienic education upon the hygienic practices of a group of individuals and families In his study, "Hygiene means all the knowledge, practices, and environmental conditions which are under the control of the family and which tend to increase health"[4] To determine the progress in hygiene, it was necessary to measure change in hygiene in the direction desired by the groups undergoing it. Observe that there is no arbitrary or absolute goal to be attained The goal is one determined by the opinions of the people. The scale to measure hygienic change was, however, constructed in an objective manner and its reliability and validity were determined. A preliminary list of 270 questions was scored on a 1000-point scale with the assistance of nine competent judges who allocated the weights From this list, a shorter form of 77 significant questions was derived[5] Sources of error were examined and shown

[3] Publications of the Faculty of Arts and Sciences, *Social Science Series,* No. 7, American University of Beirut, 1934, pp 336 ff

[4] *Ibid*, p xii

[5] *Ibid*, p 57.

not to exist as between samples, different informants in one family, different interviewers, and different scorers Reliability coefficients of + .91 on informants and + .94 on interviewers were found. Validity coefficients of + .76 were obtained with indices of mortality or "survival," of — .53 with morbidity, and of + .65 with longevity.[6]

The scale to measure hygiene having been standardized, the next problem was to select villages to serve as controls on the experimental village that received instruction. Three Arab villages were selected as controls although only one survived the test as an adequate control. These villages were located so that there was little likelihood of hygienic practices spreading to them from the experimental village of Jib Ramli. They were selected also with reference to their similarity both to the experimental village and among themselves on nine different factors which, if variable, might themselves explain the final effect, rather than the final effect being the result of the program of instruction in hygiene The controlled variables were. geographic, demographic, historical, economic, religious, domestic, educational, recreational, and sanitary conditions The extent to which these controls were quantitatively established is not clear.

In 1931, all four villages were scored on the hygiene scale. Then for two years, that is, from 1931 to 1933, the experimental village of Jib Ramli received systematic instruction by visits from the traveling clinic. The control villages received no such instruction and were by geographic location supposed to be isolated from any influence spreading from the clinic. In 1933, all four villages were again measured on the same hygiene scale and their scores compared. It was found that the experimental village increased its score from 253 points in 1931 to 304 points in 1933, a gain of 51 points, at the close of the period of hygiene instruction. The control village also increased its score, from 241 points in 1931 to 286 points in 1933, a gain of 45 points without visible instruction in hygiene Does the fact that the experimental village gained six points more than the control village prove that the effects of a hygiene program are less marked than is generally believed? Dodd makes no such claim. He concludes that the experiment may have been vitiated by incomplete isolation of the control village from the influences of

[6] *Ibid*, p 82

the clinic.[7] Our opinion is that an additional weakness in the experiment was the inability to obtain adequate control of the nine factors supposedly held constant

This example of the application of the experimental method to the study of a socio-educational problem is valuable in three ways: first, it is an admirable demonstration of the technique of constructing a scale of measurement in a difficult field of individual behavior; second, it is a clean-cut demonstration of the pattern of an experimental setup; and third, the results are allowed to speak for themselves, with critical comments by the experimenter on the weaknesses of his experiment. To prove the relationship between instruction in hygiene and progress in hygiene, similar experiments are required. In other words, a sociological experiment demonstrates the existence of a postulated relationship only after it has been repeated a number of times and its results have been verified The significance of this statement should not be lost. Note that we have said, "after it has been repeated a number of times and its results have been verified" Reproducibility and verifiability are qualities of scientific observation that the experimental method supplies to the sociologist who is patient enough to do the hard work extending over months or years, and willing to stay by his research until he secures results that can have wide acceptance.

But to return to the Dodd experiment, it is evident that the crucial problem was that of controlling the variables in the situation under investigation. Since the heart of the experimental method is the observation of the relationship of two variables when all others are controlled, it is evident that failure to control these other variables vitiates or at least obscures the results. In 1931, we published a paper, "The Problem of Controls in Experimental Sociology,"[8] in which the various techniques of control then used were analyzed on the basis of several published experimental studies Since the publication of the Dodd experiment, the problem of controls again comes to the front as a crucial problem which, if unsolved, is likely to hold back the application of experimental method to the study of social relationships

[7] *Ibid*, p xiv

[8] *Journal of Educational Sociology*, May, 1931

The Social Effects of Public Housing in Minneapolis

Is the condition of a slum family improved by rehousing in a model public housing project? An affirmative answer to this question is assumed as the justification for the expenditure of millions of dollars. Is there any proof of this assumption aside from common-sense expectation?

This study is an effort to measure the effects of good housing upon former slum families rehoused in Sumner Field Homes of Minneapolis, originally a project of the Housing Division of the PWA and since 1937 under the management of the FPHA.

The most interesting findings of this study are. (1) No significant change in morale or in general adjustment was found in 1940 as compared to 1939, either for the 44 experimental families resident in the project, or for the control group of 38 families residing in the slum; (2) both the resident and control groups gained in social participation from 1939 to 1940, but the resident families gained twice as much in absolute score as the control group; (3) both resident and control groups gained in social status from 1939 to 1940, but the residents showed a gain of greater magnitude; (4) a score made on the "condition of the furnishings of the living room" showed for the residents a striking gain but for the control group a real loss for the 12-month period; and (5) both residents and control group had improved in the percentage of families "use-crowded" in 1940 over 1939, but the gain of the residents was about three times that of the control group.

Thus the improvements in condition accrue in much larger degree to the residents of the project, and seem to justify the housing program in so far as the facts of this single study are concerned.

Before examining the evidence that supports the foregoing conclusions, it may be helpful to analyze the causal components in the complex of factors which operate in the social situation of slum dwelling compared with public housing, as well as to consider the factors chosen to measure the patterns of response to these differences in housing.

In the first place, since the purpose of this experimental study was

to discover the results of public housing, we may interpret the results as ends achieved in a means-ends schema of social reform, or we may interpret the results as the effects of multiple causation by impersonal factors. Results are ends achieved when a rule of public housing prevents room-crowding by limits placed on dwelling unit occupancy, and also when a social program results in the organization of clubs and recreational groups among public housing tenants.

On the other hand, public housing standards for interior space and lighting, reduction of noise transmission by structural standards, and regulations for safety and sanitation combine as impersonal factors in a complex of causes which result in greater measured cleanliness, orderliness, and neatness in the living room of the home, as well as a diminished percentage of use-crowding, and encourage indoor entertainment and recreation which, in turn, are expressed in higher social participation scores.

The assertion has been made that housing per se, the intrinsic housing factor, cannot be defined or measured. It may be submitted that housing per se is a combination of many factors, such as adequate space and light, reduced noise, room-crowding prevention, and provision for indoor entertainment and recreation, safety and sanitation. The response of tenants to different degrees of this pattern of intrinsic housing factors may be observed and measured by the score on the condition of living rooms, the score on social participation, and the percentage use-crowded, all three showing differences before and after public housing, as compared with continuous slum dwelling

But these specific factors, whether they be regarded as a program of means or as a causal complex, do not exhaust the list of influences at work. There are always present status factors such as the state of employment, differences in education and in income, various occupational groups, etc , and biological factors, such as the size of family and racial differences (which operate not as intrinsic biological differences but as biological variations which lead to social factors such as exclusion and segregation). Then, too, there are known but unmeasured factors such as differences in the state of health and in

habits of household economy Last of all there are many unknown factors. All of these may operate to obscure the real effect of differences in the causal pattern of intrinsic housing factors.

In summary, and in the interests of clarity of thought about this complex problem of chosen means to desired ends or of causal factors antecedent to the effect, we may list the foregoing housing influences as

M The means
  1 Occupancy rules
  2 Social program
H. Housing influences per se
  3 Space
  4 Light
  5 Noise
  6 Room-crowding
  7 Indoor provision for entertainment and recreation
  8 Safety
  9 Sanitation
S Status factors (obscuring M and H)
  10 Employment
  11 Education
  12 Income
  13 Occupation
B Biological factors
  14 Size of family
  15. Race
U (1) Unmeasured factors
  16 Health
  17. Habits of household economy
U (2) Unknown factors

Measures of Results

In studies by experimental design of the social effects of housing, we need to decide in advance of the investigation upon some measures of effect or some means of appraising the degree to which the desired end has been achieved. In this connection it is advisable to meet certain conditions as follows:

1. Scales used to measure effect should be reliable and valid instruments of observation and numerical description (see Chapter VI, pp 152–154)

2 Scales that have been used in previous studies of the same problem or a similar problem and have yielded norms on the same kind of population to be studied are valuable for comparisons, and should be chosen for use.

3. These scales should, of course, measure traits or responses that will be accepted as indices of effect or as indices of the degree to which the desired end has been achieved The scales should, in short, measure traits or responses that are relevant to the problem.

4. Such scales will be even more useful if we have prior knowledge of whether they are correlated with certain known causal variables —variables other than the complex of factors which have been selected as the means to the end desired, or the assumed cause of the effect to be studied—because this facilitates control of these disturbing factors by matching on measurements (see Appendices A and B).

5 Finally, the scales will be still more serviceable if it can be shown that they measure a *bundle* of interrelated traits that emerge as a *pattern of response* to the particular treatment or program which is the causal complex to be evaluated in terms of its results.

All these conditions are met in some degree for an experimental study of the social effects of housing if the three following measures are used. scale on the condition of the living room (see pp. 76–77, also Appendix A), percentage of dwelling units that are use-crowded, and the social participation scale (see Appendix B). The evidence for this assertion will now be considered.

1 The scale on *the condition of the living room* of a dwelling unit has shown reliability coefficients as high as $r = 95$, despite its appearance of subjectivity of estimate, and validity coefficients of bis $r = .45$ with use-crowding, and of $r = .52$ with the deficiency rating obtained by use of the *Appraisal Form* of the Committee on Hygiene of Housing (condition 1); it is a scale that has been used in several studies of many low-income families living in slum areas of the Twin Cities (condition 2); it supplies an objective measure

of such qualitative traits as cleanliness, orderliness, and condition of repair of articles and furnishings in a living room (condition 3), it shows different mean scores for different income classes and different occupational groups in an urban community, and rises from 3.9 for unskilled and semi-skilled occupational groups to 6 4 for managerial and professional groups, of a representative sample of an urban community, so that, by matching on income and occupation, that part of its variation not due to housing, but due to status factors, may be partially controlled (condition 4); and finally, it has been found related to use-crowding by bis. $r = 45$ Scores on the condition of the living room of a dwelling are obtained from part II of the *Social Status Scale* cited below.

2. The percentage of dwelling units that are *use-crowded* is a measure of the multiple use of the rooms of a dwelling. When a bedroom (BR) is used also as a dining room (DR), or a kitchen (K) is used also as a living room (LR), the situation is called use-crowded The following multiple uses of rooms originally designed to serve only one function represent degrees of use-crowding with appropriate penalties expressed numerically for purposes of measurement. LR and DR (— 6); LR and K (— 9), LR and BR (— 12) or DR and K (— 12); LR and BR and DR and K (— 15). Such situational uses distort and confuse the functions of living. Thus the incidence and degree of use-crowding is taken as an index of undesirable housing and is an index quite different from, but related to, room-crowding, or the number of persons per room in the dwelling unit Use-crowding is a sensitive measure of housing adequacy.

3. *Social participation,* or the number of clubs, formal social groups, organizations, or societies, that the persons in the family are active in, or members of, is related to provisions for indoor entertainment of visitors and to the recreation of individuals in the home. Social participation has been measured by a simple scale which obtains a total score on any person as follows. one point counted for each group membership, two points counted for each group that is attended, three points counted for each group to which some financial contribution is made; four points counted for each committee membership held; and five points counted for each of-

ficial position held, such as secretary, treasurer, president, etc This social participation scale has shown reliability coefficients of from $r = .88$ to $r = 95$, and validity coefficients of from $r = .52$ to $r = 62$ (condition 1), it has been used to measure the social activities in formal groups of the community in urban areas of both low social status and high social status, and tentative norms are available to show the mean scores obtained for different income classes and occupational groups (see Appendix B), and for both sexes (condition 2); social participation is relevant to housing since most public housing programs encourage clubs and social activities (condition 3); social participation scores have been found to be correlated with income, $r = 36$ to 49, with occupational group, $r = 31$ to 56, and with years of formal education, $r = 33$ to 44 For these reasons it is evident that by matching on education, income, and occupation, that part of any change in social participation score due to such non-housing factors as education, income, and occupation may be partially controlled (condition 4), finally, an attractive living room, not used also for eating, or cooking, or sleeping purposes, is a place in which club meetings may be comfortably held, and so social participation in the home may be related to measures of use-crowding and the condition of the living room.

Since the three measures just described were shown to be interrelated in ways expressed as correlation coefficients or by other empirical evidence, they satisfy condition 5 Furthermore, the use of the multiple critical ratio, which combines critical ratios on each of the separate measures, yields results that have high statistical significance for the results in the housing study herein described—additional evidence that condition 5 has been met.

The study was planned in 1938 to test the hypothesis the rehousing of slum families in a public housing project results in improvement of the living conditions and the social life of these families. Sumner Field Homes was selected as the test case. In an earlier study of 1935–1936, we reported on the immediate effects of slum clearance and temporary rehousing of 171 slum families[9] The pres-

[9] F Stuart Chapin, "The Effects of Slum Clearance and Re-housing on Families and Community Relationships in Minneapolis," *American Journal of Sociology*. March, 1938, pp 744–763

ent study is, therefore, a follow-up conducted upon a more systematic and experimental procedure. To test the hypothesis of improvement, we selected 108 project families (1939) as the experimental group and 131 families in slum neighborhoods as the control group.

The experimental group of resident families were those admitted to the project after December 16, 1938. The families in the control group were living in the slum and were chosen from the "waiting list," i.e., from the group of applicants fully investigated by the USHA agents but not immediately accepted as residents because they lived in poor housing not definitely sub-standard, or their income was uncertain, or there was some question of economic or social stability. They remained as eligible rejects or deferred cases for later reconsideration provided subsequent applicants did not meet the requirements in sufficient numbers to fill up the project. There were about 603 families in the waiting list For the reasons given, they were a group comparable to residents The control group of slum families was 21 3 per cent larger than the experimental group of residents to allow for shrinkage due to moving away, refusals, or other reasons

How can we measure the effects of good housing? Are residents of the project better adjusted than slum residents? The attempt to measure the effects of good housing utilized four sociometric scales that have been applied successfully in other previous studies a slum family study in Minneapolis in 1935–1936,[10] and a WPA relief study in St Paul in 1939.[11]

The scales measure· (1) *morale,* or the degree that the individual *feels* competent to cope with the future and to achieve his desired goals, (2) *general adjustment,* or the *feelings* about his relationship to other persons, toward present or future social conditions, and toward present social institutions; (3) *social participation,* or the degree to which an individual *actually* engages in the organized

[10] *Ibid*

[11] F Stuart Chapin and Julius A Jahn, "The Advantages of Work Relief over Direct Relief in Maintaining Morale in St Paul in 1939," *American Journal of Sociology,* July, 1940, pp 13–22 See also chap. III, pp 41–50

activities of his community in terms of membership, attendance, contributions, committees, and offices; and (4) *social status,* or the position the family occupies with reference to the average prevailing household possessions of other families in the community.

Interviewing of residents and non-residents began in February, 1939, and continued intermittently through July, 1939, when a total of 239 had been interviewed, 108 residents and 131 non-residents A group of 12 interviewers, graduate students in sociology and social work at the University of Minnesota, made the interviews. Only two were paid, the remainder were volunteers. The visitors were instructed in a group meeting and each was provided with sheets of typed directions before going into the field Entree to the families was obtained by the visitor's stating that he was collecting information about people's opinions as part of a wider study being made under the direction of a university scientist. No mention was made of any connection of this study with the USHA In this way it was felt that a more spontaneous response would be obtained. The interview furnished the following data.

Minnesota Survey of Opinions, two sheets with 31 questions about the individual's attitudes, to be filled in by the subject After the interview, the *morale score* and the *general adjustment score* may be extracted from the subject's marked response by a simple system of weighting and scoring It takes the subject from 20 to 30 minutes to fill this in [12]

Social Participation Scale, one sheet for entries on each group affiliation of subject recorded in five entries under five columns by the visitor in reply to questions answered by the subject It takes 10 or 15 minutes to fill in the subject's answers [13]

Social Status Scale, one sheet containing 21 entries filled in as observations made by the visitor, with perhaps one or two non-inquisitorial questions Can be completed in 5 minutes' observation [13]

[12] These scales and their norms will be found in E A Rundquist and R F Sletto, *Personality in the Depression,* University of Minnesota Press, 1936

[13] These scales and their norms will be found in F S Chapin, *Contemporary American Institutions,* Harper & Brothers, 1935, and W C Brown, 1946, pp 373–397, and F S Chapin, "Social Participation and Social Intelligence," *American Sociological Review,* April, 1939, pp. 157–166

The flow chart illustrates the actual shrinkage from the initial group of 108 resident families to the final group of 44 resident families, and from the initial group of 131 slum families not resident in the project (called the control group) to the final group of 38 families At each step in the study, the elimination of families is shown in the flow chart with the reason for this.

The 103 resident families and the 88 non-resident families that were interviewed in 1939 were matched on the following factors·

1 Race or cultural class of husband (Negro, Jew, mixed white)
2 Employment of husband (private, unemployed, OAA, WPA)
3. Occupational class of husband (I-professional, II-managerial, III-clerical, etc., using the Minnesota Rating Scale of occupations)
4. Number of persons in the family (2, 2–3, 3–5, etc)
5. Income of the family ($690–814, 815–939, etc)

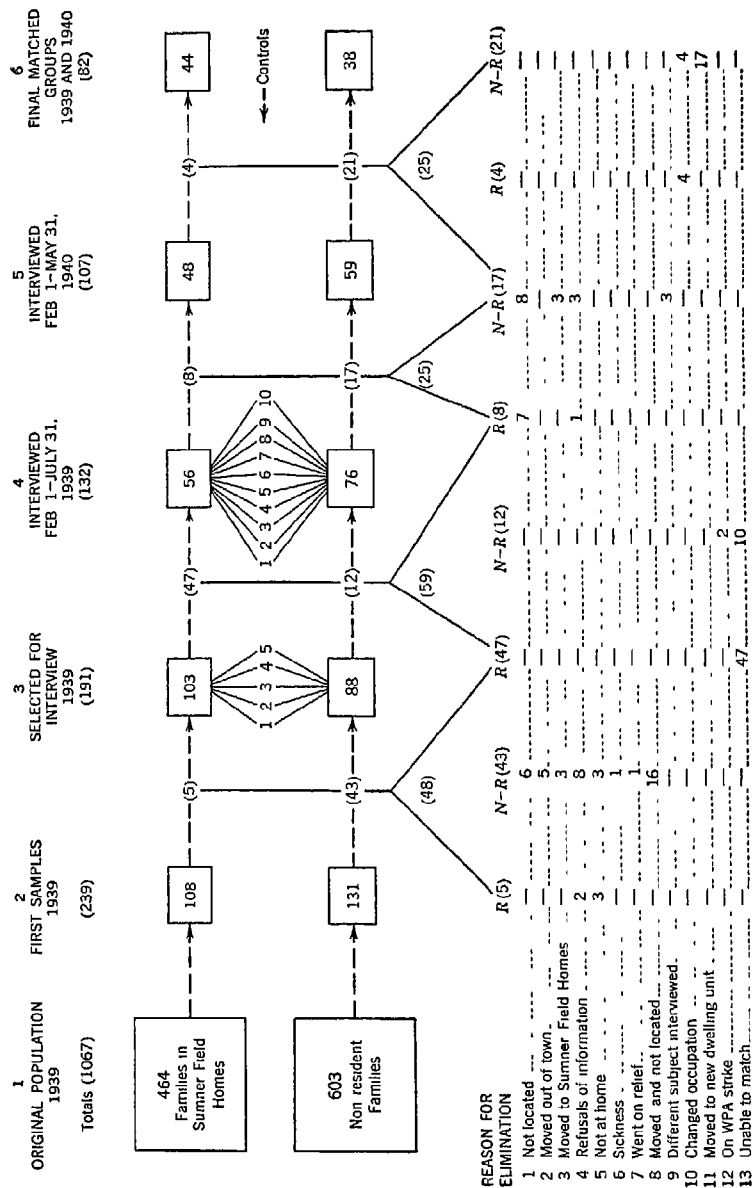
When so matched, the results of interviewing to obtain scores on morale and on general adjustment, as well as scores on social participation and social status, showed the two groups to be very much alike In fact, none of the critical ratios of the absolute differences in scores were statistically significant and in all cases were — 1 01 or less. This result establishes the fact that the initial experimental group and the initial control group matched on five factors began the experiment in 1939 (visiting was from February 1 to July 31) with a common base or zero point from which to measure change or gains.

Five additional matching factors were then added because it was found that the responses on the morale and general adjustment scales were made chiefly by housewives These five factors were:

6 Race or cultural class of wife
7 Employment of wife
8 Occupational class of wife
9 Age of wife (16–20, 21–30, etc)
10 Years education of wife (1–4, 5–8, etc)

This process eliminated 47 cases from the experimental group of residents, and 12 cases from the control group of non-residents for the reasons shown on the flow chart. This brought us to the end

FIGURE 4. Flow Chart of Effects of Good Housing in Minneapolis, 1939–1940

REASON FOR ELIMINATION	R(5)	N–R(43)	R(47)	N–R(12)	R(8)	N–R(17)	R(4)	N–R(21)
1 Not located	—	6	—	—	7	8	—	—
2 Moved out of town	—	5	—	—	—	—	—	—
3 Moved to Sumner Field Homes	—	3	—	—	—	3	—	—
4 Refusals of information	2	8	—	—	1	3	—	—
5 Not at home	3	3	—	—	—	—	—	—
6 Sickness	—	1	—	—	—	—	—	—
7 Went on relief	—	1	—	—	—	—	—	—
8 Moved and not located	—	16	—	—	—	—	—	—
9 Different subject interviewed	—	—	—	—	—	3	—	—
10 Changed occupation	—	—	—	—	—	—	4	4
11 Moved to new dwelling unit	—	—	—	—	—	—	—	17
12 On WPA strike	—	—	—	2	—	—	—	—
13 Unable to match	—	—	47	10	—	—	—	—

of the 1939 study with measurements on 56 cases of residents and 76 cases of non-residents or controls.

The next step was taken a year later (February 1 to May 31, 1940), when the follow-up eliminated 12 more cases from the resident group and 38 more cases from the non-resident group for the reasons listed on the flow chart. This left final groups of 44 resident families and 38 non-resident families matched on 10 factors The 38 families were *occupants of the same dwelling unit in 1939 and in 1940* This also added one more constant matching factor[14]

The mean scores were then calculated for these two matched

TABLE 5. Measured Changes Associated with Housing

Groups Compared	Means of Measures of Effect			
	Morale[a] Scores	General[a] Adjustment Scores	Social Participation Scores	Social Status Scores
Residents (N = 44)				
1939	60 1	45 0	1 73	60 5
1940	60 2	44 0	6 34	86 7
Change	0 1	−1 0	4 61	26 2
Critical ratio of change	0 12	−0 97	3 69	4 27
Non-residents (N = 38)				
1939	58 0	42 4	2 76	61 1
1940	56 6	41 2	4 87	82 2
Change	−1 4	−1 2	2 11	21 1
Critical ratio of change	−1 28	−1 34	2 88	3 82

[a] Reverse scales, hence minus change interpreted as a gain Critical ratios are t-values

groups and the 1939 values were compared with the 1940 values The differences or gains are shown in Table 5, together with the critical ratios of these changes.

It will be observed that the changes in morale and in general adjustment were very small, absolutely and relatively, and that the

[14] All things considered, there were in reality 13 constants the 10 factors enumerated above, the same dwelling unit at each date, and, since scores on *general adjustment* and *morale* remained approximately the same, they may be regarded as 2 additional control factors, although initially chosen as measures of effect

critical ratios of these changes show them to be not statistically significant (that is, less than 2). On the other hand, the measured changes in social participation and in social status were large in absolute magnitude and were statistically significant. This observation applies with special emphasis to the residents, who gained more in magnitude and with statistically significant gains.

There are two explanations of the insignificant changes in morale and in general adjustment First, when the raw scores on morale and general adjustment of Table 5 are converted into standard scores by the Rundquist-Sletto tables,[15] it appears that the morale and general adjustment of these housewives of slum families were about at the level of the normal population Since they were evidently not depressed or variant, it was to be expected that a change in residence for one year would have only slight effect Second, the morale and general adjustment scores of the experimental group in 1939 were obtained *after* occupancy of a dwelling unit in the housing project, so that if any gain had been experienced in relation to improved housing, it would have taken place earlier The *Survey of Opinions* form which yielded the scores on morale and on general adjustment was not part of the interviews conducted by agents of the USHA when making an initial investigation of applicants, since to have included this additional form would have increased the time of interview beyond the limit thought to be appropriate by the USHA, consequently, we were obliged to use this *Survey of Opinions* form in later interviews made by graduate student and social work visitors as described above However, all of the social participation and social status scores, as well as the information as to percentage of families use-crowded, were obtained as part of the initial interviews made by the USHA visitors, and include all of the 1067 families in 1939 The 1940 information on all scales was obtained by graduate students and social workers[16] Since the changes measured

[15] Rundquist and Sletto, *op cit*, pp 389–391

[16] A year is perhaps only a short time for changes in morale and in general adjustment to register The very slight gain on these measures shown by the nonresidents reflects perhaps the improvement in economic conditions and in general prosperity Data on total unduplicated public welfare case count for Minneapolis show a substantial improvement in 1940 over 1939. The index of store sales of the

on morale and general adjustment were so indeterminate, our remaining argument will be based upon the substantial changes in (a) social participation, (b) condition of furnishings in the living room, and (c) percentage of families in each group use-crowded.

Let us now turn to a systematic consideration of the factual evidence to discover how far it supports our original hypothesis. As stated on page 63, this hypothesis was "The rehousing of slum families in a public housing project results in improvement of the living conditions and the social life of these families"

This positive form of the hypothesis has certain disadvantages for social science research which are not always apparent to the casual reader of such statements, but which may be overcome by the null statement A null hypothesis is more susceptible of objective proof as true or false than is a positively stated hypothesis, because the latter usually makes a normative assertion, and the null hypothesis substitutes for the normative term "improvement," the neutral terms "difference between scores," "changes in scores," and "difference between changes in scores" Now it is, of course, true that the magnitude of these scores is ordered to some initial value-judgment; high scores on such direct scales as those that measure social status, condition of the living room, or social participation, and low scores on the reverse scales that measure general adjustment and morale, similarly express desirable conditions or attitudes. Thus we cannot avoid value-judgments[17] upon the results of a social

IX Federal Reserve Bank shows a change from 94 for the first six months of 1939 to 97 for the corresponding period of 1940 The foregoing explanation is confirmed by a follow-up study in 1942 of 21 of the original 44 families still (1942) resident in the project Visitors of the FPHA who made this study obtained scores on general adjustment and morale for 21 of the same housewives who had been interviewed in 1939 and in 1940, the other 23 families having moved out of the project since 1940 For these 21 women, the average general adjustment score changed 6 4 points, with a critical ratio of 2 39, or an indication of statistical significance at the 2-per-cent level! There was no statistically significant change in measured morale Furthermore, the mean scores on social status of the 23 moved-out cases were higher than the 21 that remained, which suggests that when a period of adequate length is taken, public housing is then associated with an increase in general adjustment score despite the fact of differences in social status

[17] There is a widely prevalent confusion in contemporary sociological theory which stems from semantic disorder. It is the idea that because the subject matter of socio-

program by merely substituting numerical expressions for the ordinary verbal expressions of the folk language But this is not the point at issue. The significant thing is that the scales which yielded these scores were standardized for reliability and validity *previously* and *independently* of the value-problem now before us; the scores do not, therefore, represent judgments influenced by the immediate desires, compulsions, or needs of the moment for proving a case, and hence they possess elements of objectivity (verifiability) which are not present in value-judgments of the folk language. Let us now transform the single and positively stated hypothesis into a series of three null hypotheses and examine the factual evidence which supports them or which may lead us to reject them

1. There are *no changes* in social participation, condition of the living room, and percentage use-crowded, if differences in composition of the experimental group and the control group are held constant[18] in respect to the 10 matching factors, race of husband, employment of husband, occupation of husband, number of persons in the family, income of family, race of wife, employment of wife, occupation of wife, age of wife, and years education of wife

The evidence: Table 6 shows in rows 3 and 8 for all columns that *changes were obtained* as a result of the experiment. Hence the evidence fails to support, or conversely it disproves, the first null hypothesis.

2. The observed changes in social participation, condition of the living room, and percentage use-crowded are *not greater than* those

---

logical study is the value-motivated behavior of human beings it becomes impossible therefore to avoid normative considerations, and hence that sociological theory must remain a social philosophy and cannot be divorced from values Such muddled thinking misses the essential points, which are these while it is true that human relations, the subject matter of most sociological research, *are* largely ordered to normative considerations, it is nevertheless important and entirely possible *to keep value-judgments out of the authentic language of scientific method in sociological research.* These distinctions are clearly indicated in our argument for the advantages of the null hypothesis in studies of human relations Although values are a proper and legitimate *subject of study,* they need not intrude into the methodological procedures of research to distort and to invalidate them (See chap VI, pp 141–142, chap. VI, pp 158–161, and chap VII, pp. 176–181)

[18] In the sense of control or regulation of extreme variations

that could occur between two groups selected by random sampling from the same population.

The evidence Table 6 shows in row 4 that all of these changes are statistically significant for the experimental group of residents, thus disproving the second null hypothesis in this respect. But for the control group of non-residents only column (1) of row 9 shows a

TABLE 6 Changes in Measures of Social Effects of Housing, 1939 to 1940

Groups Compared		Measures of Social Effect: Mean Social Participation Scores	Mean Scores on Condition of the Living Room	Percentage Use-crowded [a]	
		(1)	(2)	(3)	
Residents (N = 44)	1939	1 73	−0 20	50 00	1
	1940	6 34	+3 00	6 00	2
	Changes	+4 61	+3 20	−44 00	3
	Critical ratios	+3 69	+2 28	−4 44	4
	Multiple Critical Ratio		+6 01		5
Non-residents (N = 38)	1939	2 76	+3 50	44 70	6
	1940	4 87	+2 20	28 90	7
	Changes	+2 11	−1 30	−15 80	8
	Critical ratios	+2 88	−1 14	−1 43	9
	Multiple Critical Ratio		+1 82		10
Differences in Changes [b]		+2 50	+4 50	28 20	11
Critical Ratios		+1.90	+2 60	+4 11	12
Multiple Critical Ratio			+4 97		13

[a] In column (3) a negative sign (−) is interpreted as a gain, since it expressed a decline in percentage use-crowded

[b] Changes in favor of residents, i.e , changes of residents in excess of gains of non residents

statistically significant difference. But when the multiple critical ratio is calculated it will be observed that for a pattern of response (combining all three measures of social effect), the *pattern of change,* row 5, is 6.01 in statistical significance for the experimental group of residents, whereas for the control group of non-residents in row 10 it is only 1.82, or not statistically significant. Thus the evidence of Table 6 does not support the second null hypothesis

3. The observed differences between changes in social participation, condition of the living room, and percentage use-crowded are *not greater than* those that could occur frequently between two groups selected by random sampling from the same population.

The evidence The differences in changes between the two groups are all substantial, as shown in row 11, but taken individually they are statistically significant only for measures on condition of the living room and percentage use-crowded, columns (2) and (3) of row 12 But again, when the *pattern of change* in response (combining all three measures) is taken as the criterion of significance, we note in row 13 that the multiple critical ratio is 4.97, or of high statistical significance. Thus this evidence disproves the third null hypothesis.

Where does this leave us? How can we state the net result? It leaves us with the conclusion that all three null hypotheses are not supported by the factual evidence, and this is not a process of setting up straw men to be knocked down. Our net conclusion is that the residents did change significantly in the social pattern of their response to a change in housing from slum living to a public housing project, whereas the control group of families which remained in the same slum dwellings throughout the experimental period of one year changed only to a degree which could occur so frequently as a fluctuation of random sampling as to be insignificant of any real change in their condition.

It will be observed that one of the conditions of the first null hypothesis is the constancy of the ten matching factors. These factors were held constant throughout the period of the experiment. A further word is relevant, however, as to the procedure in matching. The matching process when carried out in strict manner involves

identical individual matching, that is, each individual in the experimental group is matched against another individual in the control group exactly similar in respect to the ten matching factors Since this rigorous process of matching [19] inevitably leads to heavy eliminations of cases that cannot be paired on all factors, we resorted to the expedient of pairing two or more from the experimental group against one case of the control group within a stated range. To put the matter in different phraseology, the families in the non-resident group were paired against the families in the resident group when one or more non-resident families had the same classification according to the list of matching factors as one or more of the resident families. As indicated, this procedure was less rigorous than identical individual matching but gave us greater freedom in the pairing process, prevented excessive elimination of cases, yielded terminal groups of larger size, and was followed by determinate results.

Final proof that the gains of the residents are *due solely to their improved housing* would require that we had listed all the community and personal influences that operated in the period studied and then controlled, by matching, all of these differences *excepting only* the fact that the resident group were in the project and the non-resident group were in the slum Obviously such a task would have been impossible to perform We did, however, control by matching ten factors of a personal and social nature, which, if not controlled, might have explained the differences eventually found With these ten factors [20] controlled or held constant throughout the experiment, we found by application of probability formulas that the differences measured could not have been due to chance in any reasonable expectation that reasonable persons would insist upon

Sociological research continually reveals the existence of configurations and patterns of several factors. One such pattern of factors discovered in this study was the occurrence together of higher social participation score with improved condition of the living room and less use-crowding. Since we have hitherto been dealing

[19] The *problem of interpreting differences or changes in the means of non-random* samples, such as the experimental group and the control group of this housing study, is *reserved for discussion in chap* VII

[20] The reader is again reminded that there were really not 10 controls but 13 controls, since 3 were added as the study proceeded See p 68, footnote 14

with these conditions in terms of scores (numerical symbols), it may be helpful to show the gross facts of observation from which these scores were derived. Tables 7, 8, 9, and 10 do this

Table 7 shows that the residents gained at every level of participa-

TABLE 7 Social Participation of Resident and Non-resident Groups

Social Participation Levels	Residents Frequency of Types of Participation		Non-residents Frequency of Types of Participation	
	1939	1940	1939	1940
None	29	16	26	13
1 Member	14	44	15	30
2. Attend	12	42	16	30
3 Contribute	13	37	13	24
4 Committee	0	5	1	2
5 Office	0	4	3	3
Total families	44	44	38	38

tion at least twice as much as the non-residents gained The question now may be asked, What kind of organizations were included in these gains? Table 8 supplies the answer It will be seen that the

TABLE 8 Organizations Participated in by Resident and Non-resident Groups

Types of Social Organizations	Residents Number of Persons Participating			Non-residents Number of Persons Participating		
	1939 (1)	1940 (2)	Diff (2–1)	1939 (1)	1940 (2)	Diff (2–1)
1. Sumner Field Association	0	13	13	0	0	0
Mothers' Club	0	4	4	0	0	0
2 Neighborhood house clubs	1	1	0	0	1	1
3 Church or Sunday school clubs	7	11	4	16	17	1
	1	2	1	1	2	1
4 Unions	0	1	1	0	1	1
5 Other	6	14	8	7	12	5
Total	15	46	31	24	33	9

greatest gains of the residents were in (1) the Sumner Field Tenants' Association and its subsidiaries, (2) Sunday school, and (3) other organizations What was the nature of these "other organizations"? Table 9 supplies the facts. It will be seen that in "other organizations," the residents gained by diversification and variety in their social contacts, probably a beneficial gain.

TABLE 9 Types of Social Organizations Included in the "Other" or Miscellaneous Classification in Table 8

Residents		Non-residents	
1939	1940	1939	1940
2 Social 2 Insurance 1 Bowling 1 Bridge	2 Bridge 1 Mahjong 1 Home Ec 1 Delta Theta Pi 1 W F B A 1 Sokol 3 P T A 1 Charity 1 Relief Corps 1 Scout 1 Citizen's	2 Social 2 Veterans 1 Lodge 1 Kindergarten Mother's Club 1 Scout	2 Veterans 1 Lodge 1 Mother's 1 Women's 1 Scout 3 Card (or Bridge) 3 P T A.
6	14	7	12

An explanation of the scores on condition of the living room and the subsequent differences or gains in these scores that were summarized in Table 6 can be obtained by examining Part II of the *Social Status Scale* [21] In spite of the apparent subjectivity of these

[21] The portion of the *Social Status Scale* referred to is as follows

PART II CONDITION OF ARTICLES IN LIVING ROOM

To provide some objective rating of qualitative attributes of the living room, such as "Aesthetic atmosphere" or "general impression," the following additional items may be noted The visitor should check the words that seem to describe the situation Some of the weights are of minus sign, and so operate as penalties to reduce the total score of the home

categories of observation, they are in fact very reliable, as has been shown by coefficients of reliability of + .72 to + 97 obtained from repeated observations of the same homes.

Table 6, which measures differences and gains in terms of the percentage of families use-crowded, may be explained by the information contained in Table 10 Here it will be seen that the 22 families

TABLE 10 Changes in Use-crowding of Resident and Non-resident Groups

Type of Use-Crowding	Residents $N = 44$		Non-residents $N = 38$	
	1939	1940	1939	1940
1. Dining room	1	0	3	1
2 Kitchen	0	0	0	0
3 Bedroom, or DR & K	21	3	14	9
4. BR & DR & K	0	0	0	1
Total	22	3	17	11

(or 50 per cent of the 44 resident families) classified as use-crowded used their living room as a dining room also in one case in 1939 and had no such double use in 1940 They used their living room as a bedroom also, or as a dining room and kitchen also, in 21 cases in 1939. But in 1940 there were only three such cases. This was a real

18 Cleanliness of room and furnishings.
 a Spotted or Stained (— 4)______
 b. Dusty (— 2)______
 c Spotless and dustless (+ 2)______
19 Orderliness of room and furnishings
 a Articles strewn about in disorder (— 2)______
 b Articles in place or in useable order (+ 2)______
20 Condition of repair of articles and furnishings
 a Broken, scratched, frayed, ripped, or torn (— 4)______
 b Articles or furnishings patched up (— 2)______
 c Articles or furnishings in good repair and well kept (+ 2)______
21 Record your general impression of good taste
 a Bizarre, clashing, inharmonious, or offensive (— 4)______
 b. Drab, monotonous, neutral, inoffensive (— 2)______
 c Attractive in a positive way, harmonious, quiet and restful (+ 2)______

gain in the functional purpose of the living room and represented less confusion of function in 1940 than in 1939 Similar analysis is shown for the non-resident group, which gained much less in these respects.

Table 8 shows for row 1, column (2), a gain in social participation for 17 persons in the experimental group of resident families and none for the control group This result is clearly an end achieved by virtue of the social program of public housing. But notice that even when these 17 are eliminated from the comparison, the entries in rows 3 to 5, inclusive, for the residents still show a net gain over the control group in the number of participating individuals, that is, 14 as compared with 9. This result may perhaps be interpreted as an effect of the complex of housing factors operating as a cause

The abiding purpose of the two experimental designs in which the time factor is a component, the projected experimental design and the ex post facto experimental design, is to discover a basis for generalization or for prediction. Ordinarily it would be sufficient to select random samples and to identify the probability model relevant to the kind of data manipulated to be able to generalize to a universe But, as we shall see in Chapter VII, the ideal condition of randomization both to obtain the control of unknown influences and to achieve a reliable basis for generalization is denied us because of obstacles inherent in the usual social situation. Hence the only course to follow to attain this basis of generalization is to repeat[22] the

[22] A later study of the social effects of public housing suffered from two defects· First, although it was made during 1942 and 1943, and is a before and after study, it did not utilize the available techniques and tools of measurement reported in the Minneapolis housing study described herein, although these results were published in December, 1940, and hence missed an opportunity to repeat an experimental design study on like subjects under like conditions, and second, it failed to avoid normative bias, a point commented upon by Bertram J Black, who says of this Newark housing study that it is " a curious mixture of analyses by social scientists and of a presentation for publicity purposes Testimonials, opinion polls, and conclusions are slanted to meet the needs of the proponents for public housing and detract from the adequacy of the study as a scientific document ," and again, "This reviewer hopes that when such analyses are made, the presentation of the findings be divorced completely from the desires of the housing authority for material to be used for publicity " (pp 91–92 in the *Journal of Housing*, June, 1945) The study here cited is that by Jay Rumney and Sara Shuman, *The Social Effects of Public Housing*

experiment under like conditions, if verification of results in several experiments ensues, we may calculate probabilities with some degree of assurance

An Experimental Study of Staff Stimulation to Social Participation and Social Adjustment[23]

The extra-curricular activities of university students are a controversial problem in the opinion of many persons. It is claimed that such organizational activities on a university campus serve as an agency of social adjustment. Reuben Hill conducted a study at the University of Wisconsin from August, 1940, to June, 1942, to ascertain by projected experimental design whether the claim of advantages in student participation in extra-curricular activities did in fact make for better social adjustment Among the students themselves there is a tendency to place higher value on such activities than upon classroom learning, and university counselors have frequently utilized social participation in their efforts to improve the personal adjustment of individual students.

The Wisconsin Union building provided a laboratory for the training of campus leaders, and the Division of Social Education, established to provide recreational services to students, was interested to discover the effect of its program of staff stimulation to participation in extra-curricular activities upon the scholastic achievement and the social adjustment of students Thus it seemed advisable to try to find an answer to certain questions of policy and social action in this situation Some eight previous studies of student extra-curricular activities tended to show the favorable effects of participation on both scholastic achievement and the social adjustments of college students[24]

To bring this program to a test by experimental design of study,

---

in Newark, N J, published by the Housing Authority of the City of Newark, November, 1944, and republished in March, 1946 This study illustrates the point that repetition of experimental study on similar subjects under like conditions with like methods is desirable if social research is to advance beyond an atomistic stage

[23] "An Experimental Study of Social Adjustment," by Reuben Hill *American Sociological Review*, October, 1944, pp 481–494

[24] *Ibid*, pp 481–482

the following working hypothesis was formulated. "Staff stimulation to participation in extracurricular activities results in improved scholastic achievement and improved social adjustment of college students"[25]

Since many factors combine to produce the effects of scholastic achievement and social adjustment, it was necessary at the outset to choose for control certain known factors which, if allowed to vary, might themselves produce the expected effect, quite apart from the program of staff stimulation by counseling, which latter was the causal factor to be measured. There were 15 factors selected for this control by matching on frequency distributions These may be grouped as follows: three biological factors (age, sex, and health), one personality factor (rating by high school principal); three scholarship factors (American Council percentile rank, percentile rank in high school, and scholastic average the first semester in college); four social participation factors (church affiliation, fraternity affiliation, social participation in high school, and social participation as freshmen in the university), two social opportunity factors (size of home community and size of high school); and two social status factors (type of residence and degree of self-support).

The population or universe sampled consisted of 1306 freshman students in the College of Letters and Sciences in 1940. From an alphabetical listing of these students every other name was taken to form the experimental group; those remaining constituted the control group Then the two groups were sorted out on one matching factor at a time "Where differences in the distributions of the two groups occurred the offending individuals were noted and every *nth* card rejected until the distributions were equalized"[26] By this procedure fewer losses from matching were suffered and the samples remained representative A total of 774 individuals was eliminated by inability to match them on the 15 control factors, leaving 266 each in the experimental group and the control group

The next problem was the selection and administration of scales

[25] *Ibid*, p 483.
[26] *Ibid*, p 484

to measure social adjustment before the experiment began and on the date of its termination. Conditions of reliability and validity[27] were met by two pencil-and-paper tests of social adjustment: J. P. Guilford's *Inventory of STDCR*, and J. J. Washburne's *Washburne Social Adjustment Inventory* (Thaspic edition).[28] The advantage of two inventories is that one served as a check upon the other. In November, 1940, these tests were administered in a series of meetings under the guise of gathering information for the junior dean's office. Two members of the staff of the Division of Social Education then began interviews of individuals in the experimental group. Care was taken to avoid reference to any specific questions in the tests. Individuals already active in student affairs were noted but given little counseling, the main effort at counseling being expended on the relatively inactive members of the experimental group. Key people in the student activity field, both faculty and student leaders, worked cooperatively with the counselors in directing individuals into student activities. When an individual expressed special interest in some field he was referred to the personnel head or student leader in that field. If undecided, a tentative program was worked out for him. In general, women students responded better than men students. In the 1941 guidance program the counselor used the student's record on the adjustment inventories for background information in interview. During this year 75 students dropped out of the experimental group and 78 from the control group, leaving continuous records on 86 in the experimental and 85 in the control when the experiment ended in June, 1942. It may be remembered that this period, especially early in 1942, was one of pressure to enter military service or wartime employment. In any event the numbers who dropped out were about equal in the two groups and in both some had left school, others failed to take the personality adjustment tests the second time or failed to return their questionnaires showing achievements in student activities.[29]

[27] See pp. 152–154.

[28] "An Experimental Study of Social Adjustment," *American Sociological Review*, October, 1944, p. 485.

[29] *Ibid.*, p. 488.

The results showed in a higher proportion of the experimental (treated) group active in student organizations and with its members participating in more activities per individual than was the case in the control group. About 400 student activities were evaluated and assigned weights so that the differences between the two groups showed up in higher mean score for the experimental group as well as in the number of responsible leadership positions held by its members.

In social adjustment measured by the Guilford *STDCR Inventory*, the experimental group students gained more in score than the controls in the depression factor (including feelings of unworthiness and guilt), the cycloid factor of strong emotional reactions and fluctuations in mood, in thinking introversion, and in Rhathymia (the happiness factor) Only in social introversion did the gain of the control group exceed the gain of the experimental group; and the average excess gain of the experimental group over the control, on the score for all factors, was 7 5 points, with a critical ratio of 1 74, not quite at the level of statistical significance.[30] For the Washburne test, the results were similar, but the differentials between gains in the two groups were somewhat smaller, with a critical ratio of 1 07.[31] Confirmation of the initial hypothesis that staff stimulation to participation in extra-curricular activities would result in improved social adjustment of college students is thus obtained by the results of these two tests, that is, the evidence supports and does not controvert the hypothesis. But since the critical ratios are small, the confirmation is not decisive.

On that part of the initial hypothesis which expected to find improvement in scholastic achievement, the results were indecisive, since the two groups remained practically constant throughout the experimental period[32] This result is not surprising since there are good reasons to believe that no *group* improvement in scholarship occurs during the latter years of college because gains in mean scholarship score are more likely to be a result of selection (elimination of poorer students) than of the learning process. It needs to be remembered always that the earlier heavy elimination in the freshman year is a selective process. Since no improvement in scholastic

[30] *Ibid*, p 489 [31] *Ibid*, p 490 [32] *Ibid*, p 492

achievement occurred, we may regard the scholastic record factor as another control, the sixteenth factor controlled in this study. Dr. Hill then revises his initial hypothesis to read

"Staff stimulation to participation in extracurricular activities makes for improved social adjustment of college students but its effect on their scholastic achievement is negligible." [33]

Since this experiment in guided social participation was conducted with college students, it was possible to select the experimental group and the control group at random That is to say, the group which received the benefit of staff stimulation to social participation did not have to be determined by need, as would have been the case in an experiment conducted in the natural community situation. For this reason an element of strength and integrity is added to the experiment which it seems difficult to attain under present circumstances with the pressure of the mores in social reform programs to relieve first the needy in the natural community. However, this comment should not be interpreted as an artificial limitation on this experiment which effectively disconnects it from the type of study reviewed in this book, and restricts application of the method used. As a matter of fact, student extra-curricular activities are extraordinarily varied and in many respects quite as free and voluntary as are the social activities in the natural community. Consequently it is the opinion of the author that Dr Hill's use of randomization to equalize the unknown factors between the experimental and the control groups is an important illustration of what can be done, and should more often be done, in experimental studies in the natural community situation

One final comment of a methodological nature may be worth making with respect to the hypothesis set up by Dr Hill. Considering the advantages of the null hypothesis, stated on page 70, it may be of interest to consider the type of reformulation that would convert Dr. Hill's positive hypothesis into null hypotheses. The appropriate null hypotheses are·

1. There are *no changes* between two randomly selected groups of university sophomores in social participation, in scholarship, and in social adjustment after 18 months of avocational counseling of

[33] *Ibid*, p 493

one group, when the initial differences in composition of the two groups with respect to 15 factors of biological nature, personality, social opportunity, scholarship, socio-economic status, and social activity are controlled by matching

The evidence disproves this hypothesis with respect to social participation and social adjustment.

2. The observed differences between changes in social participation, scholarship, and social adjustment are *not greater than* would occur as the fluctuations of sampling between two groups drawn at random from the same population.

The evidence does not disprove this hypothesis with any degree of decisiveness, but it does tend to disprove it.

Stated as null hypotheses, the results are, of course, the same, the only advantage being that the evidence of facts becomes somewhat clearer as a test that supports or refutes the hypothesis.

Delinquency Treatment in the Controlled Activity Group [34]

Juvenile delinquency has become a concern of parents, teachers, and police in many communities Successful treatment of this symptom of underlying social disorganization seldom occurs, and its prevention depends upon discovery of the causes and the eradication of these influences from the community Dr. Shulman reports an experimental study of the effects on individual delinquents of a new method of social treatment by means of the controlled activity group. This method utilizes the techniques of group work in a novel manner and the purpose of his study was to test the results experimentally.

Clinical or sociological analysis was lacking in the earlier approaches of group work as a preventive of delinquency Mass recreation and athletic competition did not succeed in curbing the problem because delinquents usually avoided supervised recreation or turned athletic competition into a new area for gang tactics Even the supervised club met with only relative success The development of projection methods of group therapy, consisting of loosely structured

[34] Harry M Shulman, "Delinquency Treatment in the Controlled Activity Group," *American Sociological Review*, June, 1945, pp 405–414

groups, including trained adult leaders and neurotic children admitted individually, combined the characteristics of admission of the usual educational group with the democratic pattern of a friendship group. But even in this case the therapeutic group was a segregated one and was not applicable to a large number of children who were delinquents or potential delinquents.

The group guidance experiment reported by Dr Shulman "conducted a new form of social group, the *controlled activity* group, set up within the structure of voluntary community recreation for normal children. Within the controlled activity group, problem and normal children mingled naturally in recreational activities"[85] The program was conducted under the auspices of the Social Research Laboratory of the City College of New York. It included a workshop and game-room program, classes in creative art, woodwork, metal and leather work, and met three sessions a week for two hours a session. Invited to this program were chronic truants, incorrigibles, serious personality problem children, and children charged with such serious offenses as arson and theft. Equal numbers of normal children were invited. Even the trained group leaders of the experiment were kept in ignorance of which were the delinquent children, and no school publicity was given to the experiment An informal group structure was maintained and democratic and highly individualized relationships were maintained between children and student leaders Such being the program, we now turn to the study of its empirical effects for three successive school semesters in three workshop centers for pupils from four public schools.

An experimental group of 155 children was selected to receive the controlled activity program, the results to be compared after three school semesters with a matched group of 155 children serving as the control group. The two groups consisted of boys 10½ to 14½ years of age, in grades 4A to 8A. Each group contained both normal children and problem children in the same proportions. The problem cases were obtained from official sources, and the non-problem

[85] *Ibid*, p. 407. For a discussion of the six criteria for the establishment of a controlled activity group, the reader may consult Dr. Shulman's original article.

cases were obtained by serial selection from class roll-books of all children who had neither been reported by teachers or administrators as problems, nor ever dropped below B in conduct throughout their entire school histories. The schools from which the children came were in under-privileged and low-rent areas.

Each problem child in the experimental group was individually matched against a problem child in the control group; and each normal child in the experimental group was matched against another normal child in the control group. Six matching factors were chosen. age, sex, race (there were many Negro children), educational achievement, and general intelligence measured by the *Otis Classification Test,* and mechanical aptitude by the Stenquist paper-and-pencil test There were 172 children eliminated as unable to match, and 8 more dropped out because of residence changes, leaving 130 for the experimental study Among these, 74 were matched on all six control factors, 38 on five, 16 on four, and 2 on only three of the matching traits.

Striking differences were revealed by the tests between the problem children and the non-problem children. Table 11 shows these differences. Incidentally it should be noted that the problem children were equally divided, 25 in the experimental and 25 in the control group; likewise the non-problem children were 40 in each

TABLE 11 Comparison of Problem Cases and Non-problem Cases[36]

Median Scores	All Problem Cases N = 50	All Non-problem Cases N = 80
	(1)	(2)
I Q	83	98
Educational quotient	78	94
Mechanical aptitude percentile rating	17	31
Differences between groups	Statistically significant on all measures	

[36] No tables appear in Dr Shulman's original report. Those used here were constructed by the author from the statistical data in his text, since the tabular forms make his results clearer.

group. Thus each group consisted of mingled behavior types: individual children who had responded to the invitation to participate in the program and similar children constituting the control group.

The results of the program of controlled activity as a means of treating or of preventing juvenile delinquency will now be considered in the light of the test evidence derived from the projected experimental design. These results are expressed in measured differences in behavior in two contrasting types of social situations: (1) in the classroom situation and (2) in the community situation. Although the present book is concerned primarily with a description of experimental designs applied to the study of social programs in the community situation, it may be of interest to consider the evi-

TABLE 12 Results of Experiment in Classroom Situation

Groups Compared	Per Cent Improved	Changes in Scores on Conduct Disorders [a]		
		Mild Items 1 to 5	Medium Items 6 to 10	Severe Items 11 to 15
	(1)	(2)	(3)	(4)
Problem cases of the experimental group [b] $N = 20$	48	−6 2	+7 7	Improved
Problem cases of the control group $N = 18$	24	+8 5	−1 5	Grew worse
Differences	24	14 7	9 2	Favor the experimental group
S D of this difference		9 13	7 0	

[a] On Haggerty-Wickman-Olson behavior rating scale

(2) Items 1–5 disinterest, cheating, tardiness, lying, defiance

(3) Items 6–10 over-activity, unpopularity with children, temper outbursts, bullying, speech difficulties

(4) Items 11–15 imaginative lying, sex offenses, stealing, truancy, obscenity

[b] Not all problem cases were available for case study

dence of changes in behavior in the classroom situation shown in Table 12

It will be observed that the improvement among problem children in the control group was only one-half the improvement of children in the experimental group which received the controlled activity program This appears in column (1). Turning to the last column, (4), it will be noted that children in the treated group showed a decrease in the type of symptoms which characterize the matured delinquent personality, namely, in sex offenses, stealing, truancy, etc., and that the symptoms of the control group problem children worsened The fact that column (3) shows an increase in conduct disorders among the problem children of the experimental group, and a decrease among problem children of the control group, is indicative of an increased emotional release arising from the emotional disturbance produced by experience in the workshops

Probably more significant for social policy and social action are the results found in the community situation of the family and

TABLE 13 Changes in Behavior in the Community Situation among Problem Cases

Periods of the Experiment	Changes in Family and Play Group Situations [a]			
	Experimental Group Problem Cases $N = 20$		Control Group Problem Cases $N = 18$	
	Medians	Range	Medians	Range
	(1)	(2)	(3)	(4)
After treatment	43 5	33–55	40 5	24–50
Before treatment	37 0	27–45	39 5	29–49
Changes	72% gained 2–17 point range of gain 9 5 mean point gain		33% gained 1–9 point range of gain 3 5 mean point gain	
	28% loss 2 mean point loss		66% loss 2 mean point loss	

[a] Based study of 13 of the 66 Baker-Traphagen items selected as a measure of behavior status

play group for the problem children who received the treatment; certainly the test of these results by experimental study is of central interest in the theme of this book. Table 13 shows quite conclusively that the problem cases contained in the experimental group gained far more and lost far less than did the problem children in the control group which received no treatment. It is unfortunate that only 20 of the 25 problem cases of the experimental group could be followed in case study, and that only 18 of the 25 problem cases in the control group could be similarly followed up. However, the results do stand in striking contrast for the cases studied, despite the fact that 24 per cent of the original 50 problem children cases were lost [37]

It is interesting to note that in this experimental study of the effects of social treatment, the measured gains were in several items highly significant in statistical meaning, whereas in the previous study of social adjustment reported by Dr. Hill the measured gains did not attain statistical significance despite the fact that they occurred in the expected direction. This does not mean, however, that the Shulman treatment was more successful than the Hill program. The Shulman program concerned cases of serious maladjustment; the Hill study was based on college sophomores, already a highly selected group in the direction of good adjustment. Thus we would expect significant gains in the former and be satisfied with much smaller gains in the latter, both results being indicative of the success of the program of social treatment respective to the quite different behavior standards of the subjects.

In consideration of the data presented, Dr Shulman says, "We may conclude this analysis of experimental findings by stating that the results were positive both in the school class-room and in the family setting for the treated group, that the treatment program appeared to release tensions and aggressions, to replace delinquent patterns by patterns of emotional instability, and to modify per-

[37] The results of the Shulman experiment stand by themselves as true only for the number of cases studied, since the original groups of 155 each, and the terminal groups of 65 each, were selective samples of those who responded to the invitation to join the controlled activity group, in contrast to the Hill experiment in which randomization was the basis of selection, thus providing a basis for generalization to the population of all sophomores

sonality disorders. The control problem group, by contrast, remained relatively static in school and in the family setting, some children improving slightly and the majority growing slightly worse in conduct."[38] These conclusions appear to be supported by the evidence, although without repetition of the experiment, under like conditions with similar subjects and like results, no generalization about the success of this type of treatment to juvenile delinquents in general can be made.

There remains to be considered the possible effect of factors not controlled by matching or of unknown factors To guard the results against the intrusion of "concomitant beneficent variations in societal experience,"[39] a study was made of the influence of the social backgrounds of both experimental and control problem groups, both before and after treatment, which if neglected might themselves explain the beneficial changes in behavior, rather than accepting the evidence of the results as due to treatment alone. In Table 14 are shown the changes in social background factors of family disorganization, defective social relationships, and improper discipline.

Comparison of columns (1) and (4), (2) and (5), (3) and (6) shows the existence of very marked similarities in the extent of social pathology and defective family relationships in family life of the surrounding community. As Shulman states it, "These facts lead to the conclusion that no measurable influences in the social environment of the two groups account for the marked improvement of the majority of the experimental group and the continued and slightly deteriorating misconduct of the majority of the control group of intensively studied cases."[40] Thus the conclusions about the effect of social treatment are strengthened, since Table 14 indicates that both groups were influenced in about equal degree by environmental influences that stemmed from the surrounding community life.

[38] Shulman, *op cit.*, p 413.

[39] It will be noted that these variations correspond to our own concept of natural recuperative social processes inherent in the community. See p 92 below

[40] Shulman, *op cit*, p 414

TABLE 14 Constancy of Background Factors in the Community

Characteristics and Trends of Family Background during the Period of Experiment	Community Background Factors					
	Experimental Group of Problem Cases $N = 20$ [a]			Control Group of Problem Cases $N = 18$ [a]		
	Family Disorganization	Defective Social Relationships	Improper Discipline	Family Disorganization	Defective Social Relationships	Improper Discipline
	(1)	(2)	(3)	(4)	(5)	(6)
1 Cumulative percentage	350	165	145	357	187	137
2 Combined factors		660			681	
3 Trend in family pathology	40			38.5		
4. Improvement [b]	40			27.5		

[a] Not all problem cases were available for case study

(1) and (4) percentage totals for items of broken homes, marital disharmony, public assistance, economic maladjustments, chronic illness, mental disease and deficiency, unethical or anti-social example in the home or neighborhood

(2) and (5) percentage totals for items of evidences of defective social relationships such as defective parental relationships, defective relationships to problem child, other defective relationships in the home, and defective community relationships

(3) and (6) percentage totals for severe discipline or lax discipline

[b] Commenting on the category of improvement during the experimental period, Shulman states on pp 413–414, "How much an improvement in family social relationships between problem child and parents is itself a cause or a product of behavior improvement, is however, unclear Twice as many instances occurred in the experimental as in the control group and contributed largely to the percentage difference "

A final comment may be made to point up a striking agreement in results that emerges from the Minneapolis housing study, the Wisconsin counseling study, and the New York study of the effects of treatment upon juvenile delinquency. It is this in all these experimental studies the factor of participation in social activities was an important element. In the Wisconsin and New York studies this

factor of social participation in group activities was associated with definitive gains or improvement in social adjustment as measured by standardized scales; and in the Minneapolis study increased social participation was associated with such other evidences of improvement in social adjustment as gains in score on condition of the living room and a decline in use-crowding.

These facts support a hypothesis previously stated by the author, namely, that there exist in the natural community social forces making for personal adjustment, and that these forces operate through and by way of social participation in the groups, clubs, and organizations of the natural community. We have called these forces natural recuperative social processes[41] which, if allowed free play, tend to build up and integrate the human personality, and operate like the processes of nature, without human design expressed in narrow means-ends schema. Physicians have long ago recognized the existence of natural recuperative processes in the human organism, and have come to rely upon hygiene, exercise, and rest as factors which may release these biological processes that re-create the organism. Likewise it seems entirely possible that in social treatment by social case work and social action programs, both consciously directed activities, perhaps too consciously directed, we have neglected to allow for the natural recuperative social processes that exist in the surrounding community. Certainly the evidence of these three experimental studies calls attention to the fact that the control group of families in the slum in Minneapolis gained in social participation during the period of the experiment, that the untreated students in the control group of the Wisconsin study also gained during the experimental period, and finally, that the gains of the control group of problem cases in the New York study were also of measurable degree. Thus it appears that disadvantaged groups of the community may experience gains without benefit of formal social treatment as this appears in case work and in social reform programs. If this hypothesis in verified in other studies it has im-

[41] F Stuart Chapin, "Social Participation and Social Intelligence," *American Sociological Review*, April, 1939, pp. 157–166, especially p 162.

plications of a very significant nature for all social treatment directed to the achievement of human adjustment.

All of these experimental studies stress the evidence of gains or improvement in individual adjustment of human beings As we pointed out on pages 70–71 in our discussion of the advantages of the null hypothesis over the positively stated hypothesis, there is an element of risk in the use of value-terms such as "gain" and "improvement," since our natural desire is to demonstrate by the test of experimental study that a program of treatment has in fact (rather than in mere opinion) achieved its purpose. To avoid the bias of desire we suggested that the criterion of success of treatment be stated in terms of changes in measured traits, but that this substitution of numerical expression for the ordinary terms of the folk language did not eliminate the factor of value. The important point in this matter is that the scales used to measure the changes assumed to be induced by treatment were scales that had been standardized for reliability and validity *prior to* the beginning of the particular experimental test in which they were used. Thus, although the magnitude of scores expresses a value, this value was previously and independently determined so that it could not bias interpretation of results. It will be noted that in all four experiments described in this chapter the measures of adjustment were scales previously and independently standardized, and this fact lends to the results an element of objectivity and integrity that is highly important in the study of human relations.

It cannot be too emphatically stated that, in the study of problems of human relations to determine by experimental design the results of social treatment or social action, desire is ever present to warp interpretation. It is natural for the human student of human relations to find what he wishes to find. For this reason it is important to avoid an appraisal of results which relies upon opinion. The evidence of facts and, wherever possible, the evidence of measurement are more objective and convincing forms of confirmation Consequently reliance upon opinion should be confined to the early stages of an experimental study, when decision is made upon which tools of

observation or measurement are to be selected to provide criteria of effect;[42] and also to the definition of the ends or objectives or the effects to result. After this stage of the study has been passed, the evidence of facts and measurements should be allowed to fall where it may, and inference should flow freely from this evidence without regard to opinion.

[42] Lundberg makes a similar point when he says, "The only value judgments which any properly trained scientist makes about his data are judgments regarding their relevance to his problem." *Harper's Magazine*, December, 1945, p. 530.

Chapter V

EX POST FACTO EXPERIMENTAL DESIGN: FROM PRESENT TO PAST

APPLICATIONS of the projected experimental design described in Chapter IV were made to trace and clarify our understanding of how some present causal complex of factors may produce an effect at a future date (see Table 2, Chapter II, page 32). Now our approach to the cause-and-effect relationship is made in reverse. In this chapter, therefore, we begin our study with a description of the *present situation as an effect* of some previously acting causal factors and attempt to trace back over an interval of time to some assumed causal complex of factors which began operating at an earlier date. This type of experimental study we have called the ex post facto experimental design[1] Since Ernest Greenwood[2] has analyzed and described the logic of this design, we shall confine our attention to three examples· (1) the relationship of public housing in New Haven to juvenile delinquency, (2) the effects of length of high school education on economic adjustment in St. Paul, and (3) our study, made in the 1940's, of the relationship of factors expressed in rentals to the tuberculosis death rate in New York City[3]

Two variations in the method of ex post facto experimental design will be described (a) in which self-comparison of the *same individuals and families* is made at a present date and at some past date (the New Haven and St Paul studies); and (b) in which comparison is made of populations in the *same areas* at a present date and some past date (the New York City study).

[1] Our first use of the concept, ex post facto experimental design, was made on p 29 of *Research Memorandum on Social Work in the Depression*, by F Stuart Chapin and Stuart A Queen, Bulletin 39, Social Science Research Council, 1937

[2] Ernest Greenwood, *Experimental Sociology a Study in Method*, King's Crown Press, 1945, *passim*

[3] We were assisted by Toimi E Kyllonen, John N Burrus, and Marvin J Taves.

The first type of comparison is more accurate than the second and hence preferable whenever possible, since in the second type of comparison we may not be able to account for the influence of mobility expressed as in-migration or out-migration and we are obliged to rely upon a somewhat doubtful assumption of similarity or constancy of populations and their environments at the two dates. Then there is another limitation inherent in the ex post facto experimental design which applies to both variations This is due to the restriction imposed upon us by the facts that *exist in the records* available to supply the data to be used When an experimental test is made of the results of some social program of treatment or social action by the projected design in contrast to the ex post facto design, a record of all relevant facts or opinions, as the study may require, is obtained by interview, as we have described these operations in Chapter IV. In short, *we obtain* any information that may be required by interview, questionnaire, or schedule. Usually the information so obtained is essentially new data hitherto not written into any official or public record of a government census, an annual report, a formal document, or a health registration. Such is the case with respect to social attitudes or behavior expressed as morale or social adjustment; as social participation in memberships, attendance, committee assignments, and official positions in formally organized groups; as well as in other qualities too numerous to repeat at this point

In ex post facto experimental design we are limited in choice of causal factors and in the selection of control factors to the information that exists in accessible records of the past behavior of the subjects studied. This is a serious limitation, but, despite this restriction on freedom of method and technique, there is in fact, as we shall see, a surprising number of available data for use.

Delinquency Before and After Admission to a New Haven Housing Development [4]

It has often been claimed that good housing will diminish juvenile delinquency among children of slum families simply because evi-

dence exists which shows the association of room-crowding, residence in deteriorated dwellings, and other housing deprivations with high rates of juvenile delinquency. But this evidence does not in itself alone prove that poor housing is a cause per se of juvenile delinquency. The first study in which any measure of success was realized to show a cause-and-effect rather than a mere associational relationship between poor housing and juvenile delinquency was made by Miss Naomi Barer in New Haven in 1944, and reported in 1945.[4]

The years 1940 to 1944 were taken as the date of departure in a study of 317 families who had lived in one of the developments of the New Haven Housing Authority for periods varying from two and one-half to four and one-half years The earlier records of these 317 families were then traced back over the period 1924 to 1940 before they had became residents of the public housing project. Then the records for the two periods were compared for incidence of juvenile delinquency.

There were 649 children in these families and about 70 per cent of the families were Negroes. During project occupancy the ages of the children were between 7 and 17 years. Records of the Juvenile Court were checked back to 1924 by the names of these children, together with their older brothers and sisters. The data for each year from 1924 to 1940 were then compared with the juvenile delinquencies recorded for these children to obtain incidence rates Table 15 shows the results for the 17 years before entrance to the project compared with the four years of residence in the housing development.

The difference in rates shown in column (3) between 3.18 for the pre-public housing period and 1.64 for the residence period is approximately three times the significant difference (0 52), and hence may be regarded as statistically significant. One might conclude from these results that the chief difference in family living between

[4] Naomi Barer, "Delinquency Before, After Admission to New Haven Housing Development," *Journal of Housing*, December, 1945–January, 1946, based upon "The Effects of Improved Housing on Health," M A. thesis in the Department of Public Health, Yale School of Medicine, 1945.

Table 15 Comparison of Recent Juvenile Delinquency Rates among Residents of the Housing Project with the Rates of the Same Group before Residence

Years	Total Number Child-Years in Families	Total Number Juvenile Delinquencies	Rate of Juvenile Delinquencies per 100 Children per Year
	(1)	(2)	(3)
1924–1940	1952	62	3 18
1940–1944	1401	23	1 64

the two periods, namely, residence in the public housing project, was the cause of the decline in juvenile delinquency rates.

But what of other factors? The author is careful at this point to inquire into the evidence of other factors. If the court record procedure had changed during the period, or if a notable general decrease in juvenile delinquency had occurred since 1940, the results would have no significance. Changes in the conditions of the experiment of this sort did not apparently occur, for there was an *increase* of 9 1 per cent in total juvenile delinquency in New Haven for the years 1940 to 1941, as compared with the period 1927 to 1940. Thus the decline in juvenile delinquencies among the project family children took place within the larger community in which a substantial increase in the juvenile delinquency rate was recorded.

Doubtless the reader will think of other factors that require study before the results of this experimental test of housing and juvenile delinquency can be taken as a basis for generalization about the effects of good housing. But the results seem to establish a strong presumption that the factor of improved housing played an important rôle and to this extent support the hypothesis that good housing may cause a decline in juvenile delinquency rather than merely be associated with a decline Obviously this promising experimental test of an important relationship needs to be repeated. The simplicity of the method commends it for wider application. If

replication with like methods upon similar subjects under similar conditions yields results that confirm these findings, we shall have taken a real step forward in our understanding of some of the factors causally associated with juvenile delinquency.

THE EFFECTS OF LENGTH OF HIGH SCHOOL EDUCATION ON ECONOMIC ADJUSTMENT IN THE COMMUNITY OF ST. PAUL[5]

The working hypothesis of this study was A greater degree of progress in high school leads to a correspondingly higher degree of economic adjustment in the community Thus, the two variable factors whose relationship is to be measured are ones that lie at the heart of educational policy Do we not spend millions of dollars annually to support high school education on the assumption that this hypothesis is a true statement of social relationship? Of course there are other purposes of high school education besides economic adjustment, but certainly economic adjustment is one of the most practical considerations.

This experiment was based upon the high school records and community experiences of 2127 boys and girls who left four St Paul high schools in the school year of 1926, as graduates or after having completed from one to three years of their high school course There were only four St Paul high schools at this date *so that the experiment concerns the whole universe and not a sample* The year 1926 was taken because it was the earliest year for which comparable records on a large number of students were available. Moreover, since the follow-up was to the year 1935, there was thus a period of nine years in which these individuals could work out economic adjustments.

Having set up the working hypothesis of the experiment, it next becomes necessary to secure measures of the two variables to be observed. The independent variable, school progress, was measured

[5] Appeared originally as part of "Design for Social Experiments," *American Sociological Review*, December, 1938 The experiment is that of Mrs Helen F Christiansen, "The Relation of School Progress, Measured in Terms of the Total Amount of School Attendance or Course Completion, to Subsequent Economic Adjustment," M A thesis, University of Minnesota Library, June, 1938

by the number of years of the high school course completed when the student left school in 1926. Of the total of 2127 boys and girls, 1130 graduated from high school in 1926 after completing four years and 997 dropped out in 1926 after having been in high school for the regular one or two or three years of the course. The measure of economic adjustment selected for the dependent variable was the percentage of shifts on jobs from 1926 to 1935 that involved no change in salary or an increase in salary as contrasted to the percentage of shifts that involved decrease in salary.

Now it is perfectly obvious that these are extremely crude measures Factors of age difference as between those who left at the end of the freshman high school year and those who remained to graduate might affect economic adjustment. Sex differences are often significant. Boys or girls from homes of higher status would have an advantage in gaining and holding employment not possessed by children from homes of lower status. Differences in the nationality of the parents would influence the chances of getting a job. The neighborhood of the home from which the boy or girl came might be a factor in economic adjustment. The intelligence or mental ability of the different individuals would exert its influence upon securing a job, holding the job, and upon promotion in rank and salary on the job This network of factors is also one of interrelationship. An American boy whose father was a successful professional man living in a restricted neighborhood would have several factors combined in his favor in comparison with the Jewish boy whose father was a clerk and lived in a deteriorated neighborhood. Since every one of these variable factors is recognized by sensible people as influencing the course of individual economic progress, the way to obviate their disturbing influence is to control them. Here, therefore, we come again to our crucial problem of controls.

In the Christiansen study each of these six factors, chronological age, sex, nationality of parents, father's occupation, neighborhood status, and mental ability (by computing high school marks because the measurement by I.Q was not available) was controlled. The chief contribution of the study was to demonstrate the comparative

value of different methods of controlling these factors. The pragmatic test of the value of different methods of control is the effect upon the closeness of relationship between the measured change in the dependent variable and the measured change in the independent variable But before we can present the results of this comparison it is necessary to describe some of the practical procedures that intervene between the formulation of the working hypothesis and the measurement of the results of observation under conditions of control.

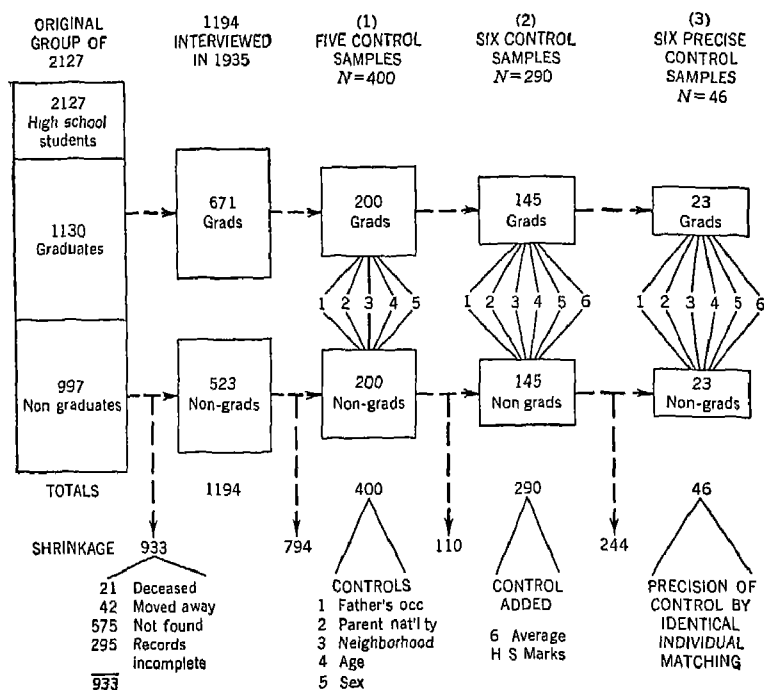
It took a full year of systematic work in home visits and interviewing to trace the 1130 graduates of 1926, and the 997 drop-outs of 1926, to their status of 1935 In this process, there was a shrinkage of 933 individuals in the total. Of the number lost, 21 were deceased, 42 had moved out of town, 575 could not be traced in the follow-up, and 295 had records so incomplete as to make comparison worthless. Thus, of the original 2127, there were located a group of 671 graduates and a group of 523 drop-outs[6] See Figure 5, flow chart

Christiansen thus had a control group of 523 drop-outs and an experimental group of 671 graduates. It was then necessary to control the six factors mentioned as potential disturbing influences on the real relationship of high school education to economic adjustment in after life. The process of gaining control began with the matching of a child from the control group and another child from the experimental group for sex and nationality of parents. This reduced the two groups to smaller groups with identical proportions in sex division and in the distribution of parental nationality At this point the control of factors by identity through individual matching had to be supplanted by control through the correspondence of frequency distributions on each factor The reason for this change was that the condition of individual identity on a factor by matching eliminated so many cases that the sample dwindled in size at an alarming rate after each new control was set.

The correspondence of frequency distributions on a given variable factor is a far less rigorous control of this variable factor than is

[6] *Ibid*, p. 7.

identity by matching individual with individual, but it is probably the most frequently used method of control in contemporary experimental studies. Its selection and application on grounds of reducing the attenuation of the sample is justified when the results of

FIGURE 5 Flow Chart

the experiment are significant. When the results are not significant and no other explanation is reasonable, then it is necessary to resort to the more expensive process of individual matching, but this means repetition of the experiment, or at least increasing the number of cases observed.

Setting the six controls reduced the final sample to a total of only 290 cases, 145 in the control group and 145 in the experimental group, a decline of 86 4 per cent from the original group of 2127

students! This is the price of observation under conditions of control The longer the list of controls and the more rigorous their method of application, the smaller the final sample At this point, some statisticians may say that we end up with a sample too small to be representative of conditions in a large original group, but let me remind them that an experiment is designed to obtain a homogeneous and "pure" sample. In fact, our sample has been purged of the very factors that made for heterogeneity in the original group and whose presence obscured the real relationship between the factors we set out to study To discover the *real* relationship between a magnet and iron, we must have "pure" iron and not iron ore that is complicated by the presence of other minerals and metals, which it would be if "representative" of the original ore Homogeneity, not representativeness, is *the* essential condition to the discovery by a single experiment of a real relationship between two factors. Consequently, if the present experiment shows that there *is* a relationship between the amount of high school education and the degree of subsequent economic adjustment, this relationship is more likely to be a *real* one than is the case under conditions in which the social situation is complicated by several uncontrolled factors Within certain limits (the limits of the given experiment), the homogeneity of the two sub-samples is more important than their representativeness of variable factors that originally obscured the relationship [7]

Finally, if we turn now to the differences in economic adjustment of the control group of drop-out students and the experimental group of graduates we find that 88 7 per cent of the graduates experienced no changes in salary or had increases in salary from 1926 to 1935, whereas 83.4 per cent of the drop-outs reported increases or no changes in salary from 1926 to 1935. Putting it the other way, only 11.3 per cent of the graduates suffered salary decreases in this period, whereas 16 6 per cent of the drop-outs suffered salary decreases. This is a small difference of 5 3 per cent.

When the length of high school education before drop-out is

[7] This is the case because random samples are seldom an available medium for social treatment in the free community situation as was shown in chap. IV in the discussion of the Minneapolis housing study This will also be considered in chap VII.

analyzed,[8] we find that 74 1 per cent who left school in 1926 at the end of one year of high school had salary increases or no changes in salary during the period 1926–1935, and of those who ended two years of high school, 85.1 per cent were adjusted economically; and 89.6 per cent of those who ended three years of high school were adjusted. Thus, in general, the longer the period of high school education, the higher the percentage of adjustment in the economic terms used as a criterion None of these differences is statistically significant as single differences The important point is that they are consistent and in the same direction. It is a matter of opinion whether small differences that are corroboratory and in the same direction are as important as differences that are large and statistically significant. Our opinion is unequivocally that small differences in the same direction may be as important as one large difference that is statistically significant. There are two reasons for this opinion. First, social phenomena are complex and not likely to show large differences because of the configurational character of the social situation. The separate factors in a social situation are usually functionally related. Second, the conventional tests of the significance of sampling are based upon the theory of random samples, and in the present stage of experimental work, as we have attempted to show, it is the terminal homogeneity and purity in the sample, rather than initial representativeness of heterogeneity, that is important in demonstrating the *real* relationship between two variable factors.

The Christiansen experiment involved the collection of other information than that about salary changes, and the analysis of this information revealed the existence of other small differences in favor of the experimental group of graduates. For example, it was found that the percentage of non-graduates or drop-outs who reported non-voting was 24 per cent as compared with the non-voting graduates, which were 14 5 per cent of the total [9] It was found that the average number of interests in activities tended to increase with the number of years' high school education one-year individuals, .50; two-year individuals, 67; three-year individuals, .63, and graduates, .91 [10] Likewise, the average number of years' additional education in-

[8] Christiansen, *op cit*, pp 64, 89. [9] *Ibid*, p 38 [10] *Ibid.*, p 41

creased with the amount of high school education.[11] One-year individuals had an average of .46 years, two-year individuals, .73 years, three-year individuals, 89 years, and graduates, 1 45 years. It will be observed that all of these differences are small but that they are predominantly in the same direction, that is, they show more favorable adjustment in direct proportion to the amount of high school education. One last measure of economic adjustment was found in the occupational class to which these young people had climbed by 1935 This rough measure of economic adjustment was also distinctly favorable to the graduate group, which had four times as many individuals in the highest occupational class as the nongraduate group Contrariwise, the graduate group had smaller numbers in the lower occupational classes.[12]

Thus the Christiansen experiment, in spite of the crudity of measuring devices used, was nevertheless successful if the argument in favor of corroboratory small differences is accepted,[13] because these small differences—in salary changes, percentage voting, average number of interests in activities, average number of additional years' education, and occupational class attained—all support the hypothesis by showing evidence of better adjustment with every increase in the amount of high school education. Doubtless there will be those who will observe that the experiment has merely proved that which we expected to be true. But justification by expectation and wishful thinking is quite a different thing from corroboration by factual evidence. On the other hand, there may be critical persons who will observe that the results of this experiment are almost too good to be true To all such, we may reply that the evidence is available for critical appraisal in the thesis, and that in any event we do not claim that one successful experiment proves a relationship to be true. The experiment must be repeated and the results verified

[11] *Ibid*, p 47

[12] *Ibid*, pp 50, 58

[13] It will be remembered that the pattern of small differences as evidence supporting the result of experimental design was summarized by the device of the multiple critical ratio in the Minneapolis housing study of chap iv It was not used in the St Paul study because of the crude character of the original observations in this study.

before the *principle* is established on factual evidence rather than on wishful opinion.

In this experiment, we have found that control of such variable factors as age, sex, nationality of parents, occupation of father, neighborhood status, and high school marks is sufficiently adequate control of the social situation to lead to definite scientific results. What would have happened if we had exercised a more rigorous control by the device of identity secured through individual pairing for every variable factor instead of relying on the rough control of correspondence between distributions in the case of the four factors of age, father's occupation, neighborhood status, and high school marks?

The answer to this question was obtained by actually pairing against each individual in the control group of drop-outs another individual from the experimental group of graduates of the same sex, identical chronological age, the same parental nationality, the same father's occupation, the same neighborhood status, and identical high school marks. This process of more exact control reduced the total from 290 to only 46, of which 23 were in the control group of drop-outs and 23 were in the experimental group of graduates. When the economic adjustment of these two contrasting groups is examined, the following results emerge: 92 per cent of the experimental group of graduates had salary increases or no change in salary from 1926 to 1935, whereas only 58 per cent of the control group of drop-outs experienced salary increases or no salary change for the period. This is a difference of 34 per cent in favor of the graduates. In comparison, the difference was only 5 3 per cent for sub-groups of the 290 cases under less rigorous controls. Thus, the labor of more exact control was amply repaid because it demonstrated, in a more decisive manner and in terms of a difference statistically significant, the superior economic adjustment of the experimental group of graduates. In this case, the real relationship was more clearly demonstrated by more rigorous controls. Of course, this further analysis might have disclosed that the real relationship was less marked, rather than more marked. Scientifically speaking, the use of controls aids us in getting at the underlying true relationship, *whatever*

that relationship may be. Whether the analysis confirms the hypothesis or disproves it is a consideration entirely irrelevant to the experimental method.

We may now summarize the conclusions from our analysis of the Christiansen experiment. (1) If there is a real relationship between two social variables and it is a substantial one, then even crude controls of other variable factors will demonstrate the existence of this relationship (2) Demonstration of the real relationship between two observed variables may be established within the limits of one experiment quite as decisively by many small differences that are in the same direction and are in agreement as by the conventional criterion of statistically significant differences. (3) When the number of cases observed is large, it repays the effort to apply control of variable factors by the method of identity through individual matching, since this procedure will demonstrate with finality (within the limitations of the given experiment) the degree of real relationship by satisfying the criterion of a statistically significant difference. To illustrate this latter point, we find that a difference of 5.3 per cent after crude control has a critical ratio of only 1.24, whereas the difference of 34.0 per cent after more rigorous control has a critical ratio of 2.90. Since such a difference would occur by chance only once in 267 (or 3.74 times in 1000) random cases, we conclude the difference is statistically significant of a real difference.[14] Furthermore, since the critical ratio of 2.90 is computed from a difference between extremely pure samples, its significance is much enhanced by this fact. The reason is that the use of the standard error of a difference as a device to determine significance is based on the theory of sampling that derives directly from chance in the throwing of identical coins. In most research, such a degree of homogeneity as found in identity is usually lacking, hence the very exacting rule of three times the standard error or one case in 370 is adopted as a guard against fluctuations caused by great heterogeneity. When, however, as in the case of the Christiansen experiment, we obtain a high degree of homogeneity by selection and control, then a critical

[14] H. E. Garrett, *Statistics in Psychology and Education*, 1930 ed., p. 134, and G. U. Yule, *An Introduction to the Theory of Statistics*, 1912 ed., p. 310

ratio of 2.0 may be accepted as quite significant of a real difference The Christiansen experiment is an ex post facto experiment and began with conditions of adjustment as they existed in 1935 and then by the method of control traced the relationship back to conditions that existed at the beginning, that is, in 1926 Wherever adequate records are available the ex post facto experiment is possible. Hence, it is very important to have good records of relevant factors to a relationship among variables that we wish to test.

We have stated our opinion that the controls in the Christiansen experiment were crude in spite of its obvious success. Let us elaborate the reasons why these controls were crude. Pairing on the factor of chronological age, for example, usually assumes that this device gives control of the degree of maturity of two individuals, and yet it is common knowledge that two persons of the same sex who are both 18 years of age may not be equally mature Within the sex factor there are doubtless differences in sex appeal which may have some marginal influence upon course grades given by a teacher of the opposite sex. In the case of such rough social status measures as nationality of parents, father's occupation, and neighborhood rating, there would be slight differences in spite of identical pairing What then is the explanation of the successful effect of such unprecise pairing? The explanation probably lies in the fact that the separate status factors tend to supplement rather than to offset one another. Similarly, there is probably a cumulative weighting of the factor of mental ability through a combination and reinforcement of influences inherent in father's occupation and high school grades The higher the father's occupation, the higher the increment of mental ability, even if slight, and this supplements the slightly higher mental ability that usually goes with higher marks in high school studies In fact, there is some support for this explanation in the results of statistical analysis of social status in which a correlation matrix of five factors of status, income, education, occupation, participation, and social status score resulted in an approach to zero for the conventional tetrad equations

Also, a later factor analysis of the same correlation matrix by Louis Guttman of the University of Minnesota, using a more elabo-

rate method, clearly demonstrated the existence of a "pure" status factor as a general factor[15] The factors other than status may also have underlying factors in common. If so, then we have an explanation of the surprisingly decisive result of using the technique of identical pairing on essentially unmeasured factors. If further research with the experimental method should support this explanation, then the problem of controls is greatly simplified whenever we have reason to believe that separately rated or separately measured factors possess a common general factor. In fact, such an explanation would supply the justification empirically, at least, for a systematic search for control factors that had a general factor in common. It would then be unnecessary to devise elaborate and expensive scales of measurement because, if the factors chosen for control were influential, they would not need to be precisely measured provided they possessed a general factor in common. Of course the requirement of more precise methods still obtains. Our conclusion merely applies to the present rough techniques of the experimental method used in empirical studies.

Review of Control by Matching for Precision of Control[16]

Many studies of social relationships qualify their conclusions by the statement: "Other things being equal," so and so takes place. For example, we attempt by public education to equalize opportunity. But how far is this end accomplished when the children of families whose parents are of foreign birth or extraction, whose father's occupation is unskilled, or who come from blighted neighborhoods drop out of school and fail to complete more often than those with greater advantages of social status? Evidently, in studies of the effect upon individual adjustment of the number of years of formal education, selective factors are at work. Hence it is too often true that other things are not equal

In this study an attempt was made to hold constant six factors

[15] Louis Guttman, "A Revision of Chapin's Social Status Scale," *American Sociological Review*, June, 1942, pp 362–369

[16] This section appeared originally in an article by F Stuart Chapin, "A Study of Social Adjustment Using the Technique of Analysis by Selective Control," in *Social Forces*, May, 1940, pp 476–487.

that would influence eventual economic adjustment if they were allowed to vary. We can therefore say we have measured the effect upon an individual child's adjustment in the community of different durations of high school education, "six other things being equal." Thus, we have tried to face squarely the limitations set by the usual phrase, "other things being equal," because *we have made approximately equal* six factors—father's occupation, parents' nationality, neighborhood status, sex, age in years, and average high school grades When these factors are not held constant, they may in and of themselves explain differences in eventual economic adjustment that might really be due to differences in length of exposure to the opportunities of high school education. By holding these factors constant as between an experimental group of girls and boys who graduated from a four years' high school course and a control group who left school before graduation, we even up the extraneous factors and bring into relief the two things we wish to measure, the amount of high school education and economic adjustment.

Figure 5 shows in graphic form the shrinkage of cases from the original 2127 to the final 46. This elimination of 2081 cases, or 97 8 per cent, illustrates two difficulties of all such studies. first, the large proportion of cases, 43 8 per cent, or 933, not found for interview in 1935 or inaccessible because of moving away, incomplete records, or death, and second, the large proportion of cases, 54 per cent, or 1148, eliminated in the process of matching on six controls. Stouffer[17] has

[17] S A Stouffer and T S Lazarsfeld, *The Family in the Depression*, 1937, Social Science Research Council, Bulletin 39, p 175 and footnote of p 176 Matching by manual manipulation (hand manipulation), i e, the sorting out of cases from the two compared groups into classes of a determined range, is a tedious procedure and results in the elimination of many cases owing to inability to match them This difficulty may be avoided by use of the formula devised by Wilks and by Guttman. See S S Wilks, "The Standard Error of the Means of 'Matched Samples,'" *Journal of Educational Psychology*, 1931, pp 205–208, and "On the Distribution of Statistics in Samples from a Normal Population of Two Variables in Matched Sampling of One Variable," *Metron*, No 3–4, 1932, also Louis Guttman, "On Uses of the Critical Ratio A Research Paper," for the M A degree, University of Minnesota, June, 1939 (unpublished), especially Section 12, "Differences between Matched Groups" (pp 29–34), and Section 13, "The Special Case of Matched Individuals" (pp 35–36) There are empirical advantages, however, that accrue to the student from manual

devised a method of analysis for dealing with cases not found, unknown cases, or refusals, which estimates what could have been the effect if the unknown cases had shown a certain amount of bias. Since the Lazarsfeld-Stouffer correction is based on the implicit assumption of a representative sampling procedure and the present study is one utilizing controls to obtain pure or homogeneous samples, we have not attempted this correction device but confine our analysis to the 1194 cases interviewed in 1935.

The purpose of the Christiansen study was to isolate the presumed cause-and-effect relationship between length of exposure to high school education as a cause (from 1922 to 1926) and economic adjustment as the effect (as found in 1935). To extricate the relationship between two factors from the usual complex of influences, it was necessary to control by matching as many factors as possible. As we have indicated, six factors which, if uncontrolled, might themselves explain the relationship were held constant as far as this could be done, considering the fact that each was essentially an unmeasured factor derived from admittedly inadequate records.

These factors may be listed in two classes. The first consists of

---

manipulation, despite its limitations, for by handling the cases he gets to know them more thoroughly Furthermore, the method of manual matching is easier to explain to the non-mathematical consumer of research results than is matching by symbolic manipulation in mathematical formulas In a somewhat different connection, Ezekiel claims an advantage for freehand curves in his graphic short-cut method for multiple curvilinear regression, and states on p 241 of his *Methods of Correlation Analysis* (John Wiley, 1930), " . In fact, the emphasis on critical study of the individual observations is a point in favor of the graphic short-cut method. It forces the researcher to know his data more thoroughly, and so to exercise thought and care in working out relations and in interpreting their significance."

In the computation of the standard error of the difference between means of matched groups allowance can be made for the fact that the process of matching introduced a degree of correlation Owing to the crude character of the matching factors used in the studies herein described, no correction was introduced for this fact If the correlation is substantial and positive (+), the size of the standard errors would be reduced, thus increasing the size of the critical ratios and increasing the statistical significance of the obtained differences Thus without this correction the results are conservative and err on the side of safety For a discussion of this matter, see C C Peters and W R Van Voorhis, *Statistical Procedures and Their Mathematical Bases*, McGraw-Hill, 1940, pp 162–167, and E F Lindquist, *Statistical Analysis in Educational Research*, Houghton Mifflin, 1940, pp 58–59.

factors of social status, such as: (1) father's occupational class[18]; (2) parents' nationality; (3) neighborhood status.[19] The second class consists of factors of individual difference, such as (4) chronological age, (5) sex; and (6) average of all high school marks[20]

The test of the effectiveness of these six controls may rest upon the results of applying them. Certainly, as we shall show, the use of these controls did in the third set of samples yield a much more decisive result in favor of the graduates than was found in the first crude control sample This result indicates that even these few controls were effective in equalizing individual factors and the social status influences favorable at the outset to graduates and unfavorable to non-graduates We are not, however, here concerned with the problem of causation so much as with an analysis of the technique of using rough non-measured factors and non-measured criteria of effect.

Before passing on to an analysis of the effect on selective criteria of adjustment of the six factors of control, it should be noted that it was not possible to control four additional factors which should be controlled in any repetition of the experiment. We refer to (1) physical health; (2) number of broken homes, (3) exact money income; and (4) persistence The last factor is quite intangible, although some studies[21] seem to suggest the possibility of its measurement and it may have an important bearing upon educational achievement and economic adjustment. No doubt other equally important factors of control will occur to the reader, but it should be

[18] Converted into numerical weights in which Class I = 7 points, Class II = 6 points, etc For composition of the classes, see the Barr-Taussig *Scale of Occupations* in Terman's *Genetic Studies of Genius*, 1925, pp 66–72.

[19] Ratings of areas of the city of St Paul by the St Paul City Planning Engineer on six classes—A, B, C, D, E, and F—converted into arbitrary numerical equivalents A = 6, B = 5, etc

[20] Taken as indication of intelligence in the absence of information about the I Q and grouped in five classes—A, B, C, D, and E—converted into arbitrary numerical equivalents: A = 93–100, B = 86–92, C = 80–85, D = 75–79, and E = below 74

[21] E G Ryans, "The Major Observable Dimensions of Behavior," *Journal of General Psychology*, July, 1938, pp 65–96 Also, "The Meaning of Persistence," *ibid*, pp 79–96, also, "An Experimental Attempt to Analyze Persistent Behavior," *ibid*, pp 333–371.

remembered that this experiment was limited to accessible crude data and is a test of what can be done with these rather than an example of a more ideally scientific study.

Turning now to the question of effect, we are again limited by the availability of criteria of economic adjustment or community adjustment that may be obtained from simple interviews conducted by untrained E.R.A. visitors. With such interviewers, it is not possible to use sociometric scales to measure morale, social participation, or other standardized criteria of adjustment. The criteria of community adjustment used initially were five in number—(1) the number of job changes from 1926 to 1935 which involved salary increases or no change in salary, a probable indication of feelings of security; (2) percentage voting, an indication of interest in political activity and of some confidence in popular government as a device to attain and to maintain social security and adjustment, (3) occupational class attained, an indication of adjustment by climbing the occupational ladder because of the association between social status and individual security; (4) additional years of education after leaving high school, and (5) the number of personal interests in diversified activity. It should be evident that these criteria, although rough, are all matters of fact that could be ascertained by unskilled interviewers.

The need of controls to equalize educational opportunity in any study that purports to isolate cause-and-effect relationship may be shown by the analysis of one example—father's occupational class—since such an analysis will illustrate the point for all factors. Table 16 shows that the graduates had a decided initial advantage in social status because 19.9 per cent of this group had fathers who were professional men and members of the managerial class, compared with only 4.41 per cent of the non-graduates. Similarly, only 10.3 per cent of the graduates had fathers who were laborers or unemployed, whereas 22.37 per cent of the non-graduates had fathers in these classes. To equalize these occupational influences, we present Table 17. Here it will be seen, reading from left to right, that samples of 400 cases and samples of 46 cases—columns (3), (4) and columns (5), (6) respectively—show progressive equalization of the initial

TABLE 16 Evidence of Advantage of Graduates

Father's Occupational Class[a]	1194 Interviewed in 1935			
	Non-Graduates		Graduates	
	Number	Per Cent	Number	Per Cent
I Professional	3	0 58	66	9 9
II Managerial	20	3 83	68	10 0
III Clerical	182	34 79	275	41 2
IV Skilled operatives	162	30 98	169	25 0
V Semi-skilled	39	7 45	24	3 6
VI Laborers	14	2 68	3	0 4
VII Unemployed	103	19 69	66	9 9
Totals	523	100 00	671	100 00

[a] Based on Barr-Taussig scale in Terman's *Genetic Studies of Genius*, pp 66–72, with unemployed added as group VII

TABLE 17 How Cases Are Eliminated When Factors Are Controlled by Matching

Father's Occupational Class	1194 Interviewed in 1935				400 with 5 Factors Controlled				46 with 6 Factors on Precision Control				46 Random Sample			
	523 Non-Graduates (1)		671 Graduates (2)		200 Non-Graduates (3)		200 Graduates (4)		23 Non-Graduates (5)		23 Graduates (6)		23[a] Non-Graduates (7)		23[b] Graduates (8)	
	No	%	No	%	No	%	No	%	No	%	No	%	No	%	No	%
I	3	0 58	66	9 9	1	0 5	4	2 0		0		0		0	2	8 7
II	20	3 83	68	10 0	10	5 0	18	9 0	1	4 4	1	4 4	1	4 4	2	8 7
III	182	34 79	275	41 2	94	47 0	94	47 0	12	52 1	12	52 1	13	56 4	11	47 8
IV	162	30 98	169	25 0	74	37 0	73	36 5	8	34 8	8	34 8	8	34 8	6	26 1
V	39	7 45	24	3 6	17	8 5	11	5 5	2	8 7	2	8 7		0	2	8 7
VI	14	2 68	3	0 4	4	2 0		0		0		0	1	4 4		0
VII	103	19 69	66	9 9		0		0		0		0		0		0
Totals	523	100 0	671	100 0	200	100 0	200	100 0	23	100 0	23	100 0	23	100 0	23	100 0

[a] Random sample of 23 non-graduates every 22nd case selected from an alphabetical list of 523 in column (1)

[b] Random sample of 23 graduates every 29th case selected from an alphabetical list of 671 in column (2)

Randomness of selection at intervals from an alphabetical list is established by a rank order correlation of Rho = 07 or practically zero, between alphabetical order and order on criterion 1, see Table 20

disparity. In columns (5) and (6) the factor "father's occupation" is exactly equalized as between the graduate group and the non-graduate group, and, as we shall soon see, this equalizing process due to matching on the occupational factor did, nevertheless, change the adjustment of the graduate group. But since in this process we equalized on six controls, meanwhile allowing length of high school

TABLE 18 Analysis of Process of Shrinkage Due to Matching on Controls

Father's Occupational Class	1194 Interviewed in 1935		Cases Eliminated NG = Non-Graduates G = Graduates	46 with 6 Factors Controlled	
	Non-Graduates 523	Graduates 671		Non-Graduates 23	Graduates 23
	Number	Number		Number	Number
I	3	66	3 NG, 66G		
II	20	68	19 NG, 67 G	1	1
III	182	275	170 NG, 263 G	12	12
IV	162	169	154 NG, 161 G	8	8
V	39	24	37 NG, 22 G	2	2
VI	14	3	14 NG, 3 G		
VII	103	66	103 NG, 66 G		
Totals	523	671	1148	23	23

education to vary, we may conclude that the increased adjustment of the graduates is probably due to the fact that they had two more years of high school education than the non-graduates. Table 18 is added merely to show how many cases of graduates or non-graduates were eliminated by the six precise controls in the final small samples of 23 each It shows more exactly than Figure 5 just how the cases were eliminated by father's occupational class as controls were established It also shows that the effect of controls was to truncate the distribution of these factors.

Let us now turn to the effective controls on criteria of community adjustment. Table 19 shows the differences between non-graduates and graduates on the five criteria of adjustment for which compara-

ble information was secured. It is evident that in every one of the three matched comparisons in columns (1), (2), and (3), the graduates (experimental group) had a higher percentage or a higher average score than the non-graduates.[22] To render the data

Table 19 Differences between Non-Graduates and Graduates on Five Criteria of Community Adjustment

	(1) Five Control Samples		(2) Six Control Samples		(3) Six Precise Control Samples		(4) Random Samples	
Criteria of Community Adjustment	200 Non-Graduates, Average 1 99 Yrs Ed	200 Graduates, Average 4 Yrs Ed	145 Non-Graduates, Average 2 Yrs Ed	145 Graduates, Average 4 Yrs Ed	23 Non-Graduates, Average 2 13 Yrs Ed	23 Graduates, Average 4 Yrs Ed	23 Non-Graduates, Average 1 9 Yrs Ed	23 Graduates, Average 4 Yrs Ed
1 Salary, per cent increase or no change	78 0	85 4	83 4	88 7	58 0	92 0	75 0	88 0
Index	100	109	100	106	100	158	100	117
Gain		9		6		58		17
2. Voting, per cent	76 0	85 5	75 0	88 0	82 6	86 9	87 0	78 0
Index	100	112	100	117	100	105	100	89
Gain		12		17		5		−11
3 Occupation class of student	4 41	5 41	4 36	5 07	4 69	4 95	4 34	5 13
Index	100	122	100	116	100	105	100	118
Gain		22		16		5		18
4 Average added years of education	0 695	1 450	0 69	1 40	0 87	1 52	0 82	1 22
Index	100	208	100	203	100	174	100	148
Gain		108		103		74		48
5. Average number of interests,	0 60	0 91	0 60	0 82	1 00	1 13	0 78	0 91
Index	100	151	100	136	100	113	100	116
Gain		51		36		13		16

more comparable, the percentage or score of the non-graduates was taken as 100 and the indicia of the graduates reduced to this base. Thus in column (1), 78 per cent of the 200 non-graduates (control group) had salary increases, or no change in salary from 1926 to 1935 in the aggregate job changes experienced in this interval, but of the 200 graduates (experimental group), 85 4 per cent, or 7.4 per cent more, had this favorable experience If, then, we take 78 per cent as a base equal to 100, the 85 4 per cent becomes 109,

[22] The only instance of loss on the part of graduates occurs in the random sample comparisons of column (4), row 2, which will be analyzed later

or a gain of 9 points on this index. Similarly, the mean occupational score of 4.41 as 100 gives to the score 5.41 an index value of 122, or a gain of 22 points on this index. Thus, by converting each comparison into an index, we can sum up the differences or gains of the graduate group on all of the five criteria in the form of numerical indicia. This conversion greatly facilitates comparison. It will be observed that the advantage of the graduate group ranges from 5 to 108 points, fluctuates on several criteria, and does not always yield the greatest advantage to the samples (3) in which precise control by identical individual matching was used. Furthermore, the most violent fluctuations of indicia occur on criteria 4 and 5. Obviously, these criteria are less stable. Critical appraisal of criterion 4, average years of education after leaving high school, indicates that this criterion gives automatic advantage to the graduate group, since a high school diploma is, itself, often a *sine qua non* for admission to further education. We shall, therefore, eliminate this criterion from future comparisons. Critical appraisal of criterion 5, average number of interests, diminishes confidence in this type of evidence because of the crude nature of the inquiry about interests made by untrained visitors. This criterion also will henceforth be dropped from our analysis.

Having eliminated criteria 4 and 5, we now turn to Table 20, in which we may examine the comparisons of summations of indicia on criteria 1, 2, and 3 for all samples. Table 20 is based upon the more dependable and objective criteria of economic and community adjustment: (1) percentage of jobs showing salary increase or no change; (2) percentage voting; and (3) occupational class score of subject. This table also enables us to analyze the relative effect of the number of controls and the precision of these controls in relation to a tentative norm of random samples.

Scores of 21 points on criteria 1 and 2 and of 43 points on criteria 1, 2, and 3 of column (1) are the summation of indicia after matching frequency distributions simultaneously on five controls applied Comparison of indicia sums of 23 points and 39 points in column (2) shows that the addition of one more control (now six controls in all), that is, by matching frequency distributions on average high school

marks, does not yield any significant gain or loss But now examine column (3), wherein six controls (age, sex, parents' nationality, neighborhood status, father's occupation, high school marks) are applied by the more precise method of identical individual matching At once, the sums of indicia increase in striking degree—that is, the

TABLE 20 Summation of Differences between Non-Graduates and Graduates Converted into Indicia of Criteria of Community Adjustment

Criteria of Community Adjustment Expressed as Sums of Indicia	(1) Five Control Samples	(2) Six Control Samples	(3) Six Precise Control Samples	(4) Random Samples
1 *Sum 1 and 2* 1 = Salary criterion 2 = Voting criterion	21	23	63	6
2 *Sum 1, 2, and 3* 1 = Salary criterion 2 = Voting criterion 3 = Occupational criterion	43	39	68	24

63 points on two criteria (row 1), and 68 points on three criteria (row 2), increases of 173 per cent and 74 per cent, respectively. Similarly, the effect of precision of control on comparisons with corresponding indicia of random samples, shown in column (4), is to yield a differential of 57 points on criteria 1 and 2, and of 44 points on criteria 1, 2, and 3, or differences of 950 per cent and 183 per cent respectively [23]

[23] The random samples of column (4), Table 20, were obtained as indicated in the footnote to Table 17 The purpose of taking a random sample of 23 non-graduates from the 523 interviewed and of 23 graduates from the 671 interviewed was to provide for purposes of comparison with the control samples of 23 each, two groups of equal size but representing unselected cases as a norm of the differences to expect on chance occurrence This enables us to apply useful theoretical criteria of probability as an additional test of significance to the simpler tests of mere absolute size

These facts establish the principle for this experiment, at least, that precision of control is more effective in the isolation of probable causal relationship between length of high school education and degree of economic adjustment than mere number of crude controls by matching frequency distributions. Precision of controls in sample (3) was obtained by identical individual matching, which means that for every boy in the non-graduate group of 23 cases there is another boy in the graduate group with the same father's occupational class, the same parents' nationality, from a home in the same neighborhood, of the same age, and with the same average high school marks The same matching procedure applies to girls. It is evident that these stringent conditions for equalizing differences on six controls are so severe that 244 cases are eliminated because they cannot meet one or more of these conditions.

Let us now examine more carefully some of the subtler implications of the technique used in constructing Table 20, since this one table is the chief evidence for the contention that cause-and-effect relationships have been partially isolated by means of comparatively crude controls and factual criteria.

It is evident that the differences between columns (3) and (4) of row 1 (that is to say, 63 points compared to 6 in Table 20) are accounted for by the plus 34 per cent advantage of graduates—Table 19, column (3), row 1—in the precisely controlled sample, and by the minus 9 per cent disadvantage of graduates in the random sample—Table 19, column (4), row 2 Since our argument in justification of the utility of precise controls rests on these two data, it becomes necessary to inquire whether the differences of plus 34 per cent in one case and minus 9 per cent in the other case, which are the raw differences that underlie the large differences in indicia, are statistically reliable An answer to this question is found by calculating the probability of such differences occurring by chance

Is the difference of plus 34 per cent likely to occur often in small random samples as a chance fluctuation? If so, such a difference is not significant. But exactly how often would such a difference, or a slightly larger one, occur in random sampling? It would occur

once in about 200 trials.[24] Chance occurrence is therefore rare. Hence, when it occurs in practice, the odds are about 200 to 1 against such a difference occurring by chance, and we regard its occurrence as significant. Correspondingly, in the random sample for a difference of minus 9 per cent, the chances of such a difference are 1 in 14, or about 13 against 1 of such an occurrence by chance.[25] Since, in the comparison, an occurrence of minus 9 per cent is sufficiently frequent by chance so that when it occurs in practice it cannot be regarded as significant, we may therefore conclude that the positive advantage, plus 34 per cent of the graduates in the precision sample, is a reliable difference and that the negative position, minus 9 per cent of the graduates in the random sample, is not a reliable difference. By this line of reasoning, we find the theoretical support for our contention that the differences among indicia in Table 20, columns (3) and (4), establish the utility of precise controls in this kind of analysis. Of course, we have only the evidence of one experiment. More adequate proof requires verification by repetition of this type of experiment.

Let us now examine more minutely the two distinct criteria of economic adjustment, salary change and student's present occupational class. Table 21 shows the variation and percentage, salary increases and job changes, and the percentage experiencing no change and the percentage with salary increases. It will be observed that the large samples, (1) and (2), show more decisive gradation with years of educational exposure than the precise control sample (3), in which the fluctuations are more pronounced, as is usually the case in small samples. When, however, the totals of all non-graduates, whether with one or two or three years of high school attendance, are thrown together and the percentage is recomputed

[24] The critical ratio of the difference of 34 per cent between the non-graduate group and the graduate group is 2.83 by the following formula

$$S.D._{(p_2 - p_1)} = \sqrt{\left(\frac{N_1 + N_2}{N_1 + N_2 - 2}\right)\left(\frac{p_1 q_1}{N_2} + \frac{p_2 q_2}{N_1}\right)}$$

being an adaption of Fischer's "t" formula to the standard deviation of a proportion.

[25] The critical ratio of the difference of minus 9 per cent between the non-graduate group and the graduate group of the random sample, by the same formula, is — 0.18.

for comparison with all graduates, we find that only 58 per cent had salary increases or no change in salary, compared with 92 per cent (the sum of 54 and 38) of the graduate group. Thus, although the samples (1) of 200 cases each show a regular gain in economic adjustment with each additional year of education—column (1)—

TABLE 21 Economic Adjustment Per Cent Changes in Jobs, 1926–1935, with Salary Increases, No Change, or Decreases

Left High School in 1926 After	Per Cent Salary Increases[a]			Per Cent Salary,' No Change[a]			Per Cent Salary Decreases		
	(1) 5— Control Samples $N = 400$	(2) 6— Control Samples $N = 290$	(3) 6— Control Samples $N = 46$	(1) 5— Control Samples $N = 400$	(2) 6— Control Samples $N = 290$	(3) 6— Control Samples $N = 46$	(1) 5— Control Samples $N = 400$	(2) 6— Control Samples $N = 290$	(3) 6— Control Samples $N = 46$
1 year's attendance	28 4	25 9	33 3	42 0	48 2	25 0	29 6	25 9	41 6
2 years' attendance	38 4	46 3	26 7	39 5	38 8	40 0	21 9	14 9	33 3
3 years' attendance	47 7	56 8	31 2	37 2	32 8	18 7	15 1	10 4	50 0
4 years' attendance, graduates	55 1	53 2	54 0	30 3	35 5	38 0	14 6	11 3	8 0

[a] In this table the sum 55 1 (per cent salary increases) and 30 3 (per cent no salary change), of the 200 graduates (bottom row) is 85 4 per cent, which appears in Table 19, column (1), row 1. Non-graduates added together and per cent for this group of 200 recomputed in Table 19, column (1), row 1

the dichotomous comparison for the small samples (3) yields an even more favorable test for the graduate group. This evidence justifies dependence on criterion 1, salary increase or no salary change during 1926–1935, as a test of the techniques used in this controlled analysis.

On the other hand, the criterion of student's occupation in 1935 is not as satisfactory. Table 22 shows that the occupational class differences in favor of the graduates fluctuate from 23 3 per cent for professional and managerial occupations in the 1194 original group to 21.5 per cent in samples (1), to 15 6 per cent in samples (2), and a rise to 22 1 per cent in samples (3) Thus the control of factors does not seem to enhance the effectiveness of the occupational criterion. For this reason, we conclude that row 1 of Table 20 ap-

plies the best test of the utility of this experiment and establishes more decisively than the sum of three criteria the effects of precision of control on six factors.

TABLE 22 Analysis of Percentage Differences between Non-Graduates and Graduates on Criterion of Student's Occupational Class

Occupational Class of Students	Original 1194 Interviewed in 1935			(1) Five Control Samples			(2) Six Control Samples			(3) Six Precise Control Samples		
	523 Non-Graduates	671 Graduates	Differences	200 Non-Graduates	200 Graduates	Differences	145 Non-Graduates	145 Graduates	Differences	23 Non-Graduates	23 Graduates	Differences
I	2 3	19 8	+17 5	4 0	16 0	+12 0	4 0	13 1	+9 1	4 4	17 8	+13 4
II	4 5	10 3	+5 8	3 5	13 0	+9 5	3 6	13 1	+6 5	0	8 7	+8 7
III	47 1	48 0	+0 9	46 0	49 0	+3 0	48 2	53 1	+4 8	65 2	39 1	−26 1
IV	32 1	12 0	−20 1	33 0	16 5	−16 5	29 0	15 2	−14 8	21 7	26 0	+4 3
V	4 9	2 3	−2 6	6 0	2 0	−4 0	5 5	2 1	−3 4	8 7	4 4	−4 3
VI	3 0	0 8	−2 2	4 5	1 0	−3 5	3 6	1 4	−2 2	0	4 4	+4 4
VII	6 1	6 8	+0 7	3 0	2 5	−0 5	6 1	2 0	−4 1	0	0	0
	100 0	100 0		100 0	100 0		100 0	100 0		100 0	100 0	

Since this conclusion limits the test of the experiment to only two criteria, namely (1) per cent salary increases or no changes, and (2) per cent voting, we turn finally to Table 23, and consider the probability of differences of specified size between percentages occurring by chance. Taking the percentages found in the random samples as a norm for comparison with the small samples having precision control on six factors, we may ask whether the difference of 21 per cent between the 34 per cent advantage of the graduates in the control sample and the 13 per cent advantage of graduates in the random sample is a difference or gain that frequently occurs by chance. The last column shows that the approximate critical ratio is 6 04 for a 21 per cent difference in advantage on criterion 1. Since, by probability tables,[26] the odds of such occurrence by chance

[26] H Arkin and R C Colton, *An Outline of Statistical Methods*, 1938, pp 114 and 120. The approximation critical ratio formula for a difference between gains is

$$C.R. = (\bar{v}_2 - \bar{v}_1)\sqrt{\frac{N(N-1)}{(p_1'q_1' + p_1q_1 + p_2'q_2' + p_2q_2)}}, \text{ when } N_1 = N_2.$$

This formula is adopted for use in dealing with proportions from formula (5) by Louis Guttman, in "On Uses of the Critical Ratio," p 26, a research paper for M A degree, University of Minnesota, June, 1939.

are approximately only 1 in 500,000,000 trials, we may conclude that this difference is highly significant Similarly, the approximate critical ratio for a difference of 15.3 is 4.53, with odds of 1 in over 2000 of

TABLE 23 Analysis of Probability of Occurrence of a Difference between Differences on Criteria of Salary and Voting

Criterion of Adjustment	6 Precise Control Samples		Random Samples		Differences of Differences	
	Non-Graduates	Graduates	Non-Graduates	Graduates	Difference of Per Cent	Critical Ratio
1 Salary, per cent increase or no change	58 0	92 0	75 0	88 0		
Difference		34 0		13 0	21 0	6 04 [a]
2 Voting, per cent	82 6	86 9	87 0	78 0		
Difference		4 3		−4 0	8 3	

[a] In this table we assume that the proportions and the differences found for the random samples on each criterion represent the normal differences between non-graduates and graduates Then the greater differences found for the controlled comparisons may be interpreted as differentials isolated by controlled analysis and their statistical significance tested as above

such a difference occurring by chance. Therefore, this difference is also highly significant [27]

The entire analysis may now be summarized under six main points:

1. Simple factual records may yield highly significant differences that test a cause-and-effect hypothesis, provided a large number of cases are used at the outset (2127 in this study) and a long period intervenes between cause and effect (nine years in this study).

2 The elimination of cases is very heavy in a long-time study and especially as the number and precision of controls increase.

3 Simple factual data gathered by untrained interviewers may yield significant results when the control group technique is used.

4 Making the controls precise by identical individual matching

[27] Again we call the reader's attention to the fact that in chap. VII there is a discussion of the limitations imposed by the application of the conventional probability tests derived from random samples to such cases of non-random samples as experimental and control groups used in the experimental designs of this book.

on the factors available is much more effective than mere number of controls used without precision.

5. Recently devised formulas for computing the critical ratio of differences among proportions in samples provide a test of the probability of occurrence of these differences by chance.

6. The techniques developed in this ex post facto experiment seem sufficiently successful in isolating possible cause-and-effect relationships to suggest that they provide an additional method of research in longitudinal studies of human behavior.

Rentals and Tuberculosis Death Rates in New York City [28]

In the two preceding applications of ex post facto experimental design some present effect of a social program upon individuals or families was traced back to an earlier date when this program began operating as a cause, or as a means to achieve some desired end. In both instances the *same individuals and families* were compared over the experimental period In the study now to be described an attempt was made to apply this experimental design in an analysis of assumed cause-and-effect relationships in which the units studied were not the same individuals or families but the populations residing in the *same areas* for the experimental period. Consequently this study is in human ecology and presents the results of an exploration of the possibilities of analysis of cause-and-effect relations in a field of social phenomena hitherto investigated by different methods It is hoped that the results will be regarded as sufficiently definitive to encourage further application of the method to other ecological relationships in which causal factors are postulated as antecedent to some present effect, so that our understanding of human relations may extend and deepen.

Are high standards of housing in 1930 related to lowered tuberculosis death rates in 1940? Is it possible in the use of analysis by experimental design to isolate the antecedent influence of a complex of housing factors upon the tuberculosis death rate of a later date?

[28] In this study the author acknowledges the assistance of three graduate students, Toimi E Kyllonen, John N Burrus, and Marvin J Taves, in a graduate seminar at the University of Minnesota during the year 1945–1946.

To test the possibilities of finding an answer to these questions a large city (universe of small areas) was selected in which the decline in tuberculosis death rates 1931–1940 was, on the average, not significant[29] except as a small segment of a long-time trend. Then a group of low tuberculosis death rate areas of 1940 was compared with a group of high tuberculosis death rate areas of 1940, by tracing back to 1930 for each area the magnitude of the mean monthly rentals (taken as a rough measure of a bundle of interrelated housing factors).

The results of this analysis by ex post facto experimental design are as follows (1) a decline in tuberculosis death rates of high statistical significance[30] from 1931 to 1940 was *preceded* by high standards of housing in 1930; and (2) a decline in the tuberculosis death rates of non-statistical significance[31] from 1931 to 1940 was preceded by low standards of housing in 1930. Thus the conclusion seems justified that high housing standards precede a decline in tuberculosis death rates, but that low housing standards in areas studied are not related to a decline in tuberculosis death rates.

The foregoing brief summary of results now requires detailed explanation of the sequence of operations performed to obtain them (1) a description of the universe of study and the samples drawn from it; (2) definition of high and low housing standards, (3) description of the procedures in an ex post facto experimental design; (4) statement of variations from former procedure in this design; (5) limitations and weaknesses of this study; and (6) possibilities of further application of this method of analysis of social relationships

The Universe and the Samples

There were 354 health areas in New York City for which in 1940 comparable data were recorded (1) the tuberculosis death rates per

[29] A critical ratio of change in tuberculosis death rates from a stratified random sample, 1931–1940, was 0.59

[30] The critical ratio of change in tuberculosis death rates in high rent areas was 4.6

[31] The critical ratio of change in tuberculosis death rates in low rent areas was 1.90.

100,000 of population (X_1),[82] (2) median monthly rentals (X_5); (3) percentage of population native white (X_3); and (4) percentage of population colored (X_4) These health areas are the universe from which samples for 1940 were drawn.

To obtain the greatest contrast in tuberculosis death rates the 354 health areas were put in array on these rates for 1940 and the upper quartile of areas ($N = 88$) was chosen for comparison with the lower quartile areas ($N = 88$). These two samples from the extremes of the distribution of tuberculosis death rates—shown in Table 24, column (3), rows 1 and 4—were matched by frequency distributions on percentage native white (X_3), and percentage colored (X_4). In this process of equating by frequencies, 47 health areas were eliminated from each group because they could not be matched. The remaining 41 high tuberculosis rate areas are hereinafter called the "experimental group" and the 41 low rate areas are called the "control group" (See Table 24, rows 2 and 3)

Since it was desirable to obtain a norm to which the experimental group and the control group could be compared, a stratified random sample of 42 health areas was selected from the 354 areas of New York City in 1940 as follows The stratification was by city boroughs, i e., the data for each health area having been recorded on a separate card, these cards were grouped by boroughs, then each group of cards was shuffled by one person while another took out every seventh card in each borough group; in this manner the stratified random sample mentioned above was obtained.

[82] Actually the tuberculosis rate data are the mean rates of 1931 to 1935 taken as 1931 and the mean rates of 1936 to 1940 taken as 1940 See Bellows, Drolet, Goode, and Widdemer, *Neighborhood Health Development, Health Center Districts,* Department of Health, City of New York, Handbook, Statistical Reference Data, Ten-Year Period, 1931–1940. Bellows, Drolet, Goode, and Widdemer report the existence in 1940 of 354 health areas in New York City However, they record comparable data for 1930 and 1940 for only 306 of these areas In the present study, the 354 health areas have been considered the universe, a quartile then consists of 88 of these areas These 88 cases have been selected from the 306 areas for which comparable 1930 and 1940 data are actually available The quartile here reported thus represents more than 25 per cent of the 306 areas on which the study must be based The net effect upon the results is to give a more conservative picture than is actually true, since these "quartiles" are "closer" to each other than true quartiles would be

TABLE 24 Contrasting Tuberculosis Death Rates in Contrasting Rental Areas, New York City, 1930–1940

Contrasting Areas	1930		1940	
	TB Rates	Rentals	TB Rates	Rentals
	(1)	(2)	(3)	(4)
1 Upper quartile on TB rates	115 45	$39 88	106 45	$29 88
2. Experimental group	96 96	$36 65	83 97	$28 80
3 Control group	32 00	$72 48	22 63	$48 41
4 Lower quartile on TB rates	32 11	$70 61	11 15	$46 65

Definition of High and Low Housing Standards

Median monthly rentals (X_5) for each health area were taken as a rough index of housing conditions for 1930 and 1940. Other indices of housing standards of greater validity would have been preferable to median rentals, but *it needs to be remembered that rentals were the only recorded index of housing for these areas available to us.* Thus the index of housing standards although admittedly weak would, if the results were significant despite an imperfect index, yield conclusions which would be at least conservative as compared with the probable results of a more perfect index of housing conditions.

The results of other studies (based on other areas) of the relationship between rentals and housing factors show that rentals are often correlated with tuberculosis death rates, the percentage crowded, the income of the family, etc. For example, Schmid[33] found a correlation of $r = -.532$ with tuberculosis death rates, of $r = -555$ with percentage crowded, and that this latter correlation persisted, i.e., did not drop below $-.40$, despite partialing out other

[33] Based upon 450 enumeration districts of the Twin Cities, *Social Saga of Two Cities*, 1935.

variables such as percentage needing major repairs, percentage of dwelling units without central heat, percentage of structures 35 years of age and older, percentage of one-family structures, and percentage without gas. Furthermore, he found a multiple correlation of .743 between rental and overcrowding and mean value of structure. Britten[34] found an increase in the relative prevalence of secondary tuberculosis cases for all ages under 65 years of age with increase in room-crowding, age, and income controlled. The increases were most closely associated for tuberculosis cases 15 years of age and under. *The Minnesota Income Study* of 1938-1939 reports a correlation of .678 between rentals and income. A study of the Housing Authority of the City of New Haven of 17 families with a tubercular history and resident in the public housing development showed a favorable prognosis in 15 of the 17 cases, whereas a control group of 17 families matched on significant factors showed only 9 families with favorable prognosis.[35] These findings are the basis for the assumption made in this study that rentals are a rough index of a bundle of related housing factors. We assume, therefore, that high rentals are a partial index of high housing conditions, and vice versa.

The Procedure of Ex Post Facto Experimental Design

In the present study tuberculosis death rates of 1940 were taken as the dependent or effect variable and the health areas recording these rates were studied in 1930 for variations in rentals, taken as the independent variable or complex of causal factors.[36] Obviously such a study is strictly limited by available records of present and past events or trends, since no new measures of past events can be made in the present. Hence the causal factors are likely to be far fewer than one would wish them to be and are much more limited than in the projected experimental design wherein a follow-up to some future date is contemplated. Furthermore, the controls available may

[34] *American Journal of Public Health*, February, 1942, Fig 2, based on data of the National Health Survey of 1935-1936

[35] *Journal of Housing*, December, 1945-January, 1946.

[36] This complex of antecedent causal factors consists of the combination of the unplanned consequences of many independently planned social actions, or, to state the matter briefly, economic forces.

be less numerous and precision of control less easy. Hence in the present analysis the dependent or effect variable (X_1), is the tuberculosis death rates of 1940 The independent variable we desire to test or measure, in both concomitant and prior variation with the tuberculosis death rate, as the assumed causal variable or complex of factors, is median rentals (X_5).

In the present study, the order of procedure from a chosen effect (X_1 = TB death rates) to an assumed cause (X_5 = rentals) was followed. An alternate procedure in ex post facto experimental design is from a chosen causal factor (X_5 = rentals) to an assumed effect (X_1 = TB death rates). One of the authors, Mr. Taves, tried out this second order of procedure[37] on the same New York City data. It was found that comparison of a group of low rent areas ($N = 36$) in 1930, matched by frequency distribution on percentage colored in 1930, showed in 1940 rates of tuberculosis which diminished 13.3 per cent for the low rent areas of 1930 and 20 9 per cent for the "medium" rent areas of 1930, the former decline having a critical ratio of 1 3 and the latter a critical ratio of 2.8 These results are consistent with the results of the first method, but less decisive, because the rent groups (low and medium) were of less extreme magnitudes than were the tuberculosis death rates used in the first comparison.

Now there are two other variables recorded for the New York City health areas which are known by previous studies to be related to the tuberculosis death rate; these are the percentage native white and the percentage colored in the population. Green[38] found a high relation between a measure of economic conditions (monthly rental and possession of radio and telephone) and the tuberculosis death rate for the four-year period 1928-1931 in 252 census tracts in Cleveland. For whites the tuberculosis rate ranged from 19 in high economic status areas to 127 in the lowest economic status areas, and the rates for Negroes were five times the white rates in each area.

[37] This supplementary analysis is what Greenwood (*op cit*, pp 64–68, and chap 9) calls ex post facto cause-to-effect experiments, a process of operations in reverse to the sequence followed in the three studies herein summarized

[38] *Tuberculosis and Economic Strata, Cleveland's Five-City Area, 1928–31*, the Anti-Tuberculosis League, Cleveland, 1931.

Whitney[39] found the tuberculosis rate highest where the living standards were lowest in a study of tuberculosis death rates by occupational groups in several states.

In the present study this same inverse relationship between tuberculosis death rates and economic factors (median rentals) and race (colored) was also found.

Table 24 shows that high tuberculosis death rates (rows 1 and 2) occur where mean rentals are low, and low tuberculosis death rates occur where mean rentals are high (rows 3 and 4). These results are derived from a comparison of extreme rate areas. What is the relationship for the city as a whole? Taking the stratified random sample for an answer to this question we find that the correlation between median rentals[40] and the tuberculosis death rate is $r = -308$, and between percentage native white and tuberculosis rates is $r = -930$, both results are thus consistent with former findings

Since, however, a correlation between tuberculosis death rates and median rentals, $r_{15} = -.308$, in the stratified random sample is lower than Schmid's result of $r = -.532$, one wonders whether the relationship found in the stratified random sample for New York City is masked by the larger colored population there than in the Twin Cities studied by Schmid. To test this point the partial correlation in the stratified random sample between the tuberculosis death rate and median rentals was computed, holding constant the percentage colored.[41] The effect was at once striking, since the partial correlation was $r_{15\,4} = -66$, or double the magnitude of the zero order correlation, $r_{15} = -308$.

It appears, therefore, that control of percentage colored is desirable if the housing factor (rentals) is to be partially isolated. Thus the much higher partial correlation on the effect variable[42] gives us the

[39] *Death Rates by Occupation,* based on data of the U S Census Bureau, 1930 National Tuberculosis Association, 1934

[40] When mean rentals are cited in this paper, the figure is the mean of median rentals for each health area, and correlations are based on recorded median rentals, whereas critical ratios are based on means of these medians

[41] Of the two nationality and racial factors, the percentage colored is the more important factor, since $r_{15\,4} = -66$

[42] In another study of infant mortality and median rentals for the same areas, the partial correlation from a stratified random sample increased to $r_{12\,34} = -33$, from

raison d'être for the selection of matching factors (to be held constant) which are to be used in the construction of an experimental group and a control group as we attempt to trace back from 1940[48] to 1930 with the purpose of discovering an assumed cause-and-effect relationship operating in and by the rentals of 1930.

We are now ready to consider the tabulated figures upon which the conclusions are based which were stated at the beginning of this section

Table 25 is arranged to show in descending magnitude the tuberculosis death rates of the different samples from the top to the bottom of the table. It will be observed that one effect of matching the upper quartile with the lower quartile on percentage native white and colored is to diminish the range in both tuberculosis rates and mean rentals The stratified random sample stands in a middle position (row 3) on all rates, which was to be expected It will be noted also that the critical ratios of changes in rates from 1930 to 1940 (column 5) are non-significant for the tuberculosis death rates in high rate but low rental areas (rows 1 and 2), as also for the stratified random sample (row 3); but that the changes in tuberculosis rates in the low rate areas of high rentals (rows 4 and 5) are highly significant, the critical ratios being all in excess of 4. Further examination of the high tuberculosis rate areas (row 2) shows that the rentals dropped non-significantly (less than 3 standard errors) from \$36 65 to \$28.80 in 1930–1940, whereas in the control group of low tuberculosis rate and high rent areas (row 4) there was a significant decline (4 2) in rentals from \$72 48 in 1930 to \$48 41 in 1940 The important point to note is that, despite this decline of significance in rentals for the high rent areas 1930 to 1940, the level of rentals for these areas was even at its lowest value

$r_{13} = -.24$, where X_2 = infant mortality, and subsequent analysis by ex post facto experimental design yielded definite but less convincing results than in the tuberculosis-rental study, all of which indicates the value of pre-test by partial correlation

[48] The constancy of control on percentage colored set in 1940 was maintained for the same areas in 1930, the mean percentage colored in the experimental group was 1 48 per cent in 1940 and 1 7 per cent in 1930, and in the control group was 0 78 per cent in 1940 and 0 65 per cent in 1930

TABLE 25 Comparison of All Samples for Tuberculosis Death Rates[a] per 100,000 Population and Means of Median Rentals for 1930 to 1940, New York City Health Areas

Samples	1930		1940		Critical Ratios of Changes
	TB Rates	Rentals	TB Rates	Rentals	
	(1)	(2)	(3)	(4)	(5)
1 Upper quartile, $N = 88$	115 45	$39 88	106 46	$29 88	−1 21 −8 99
2 Experimental group, $N = 41$	96 96	$36 65	83 87	$28 80	−1 90 −2 40
3 Stratified random sample, $N = 42$	55 30	$53 86	50 24	$37 54	−0 59 −5 14
4 Control group, $N = 41$	32 0	$72 48	22 63	$48 41	−4 6 −4 2
5 Lower quartile, $N = 88$	32 11	$70 61	22 15	$46 65	−5 28 −7.15
7 United States[b]	67 8	——	45 9	$27 31	—— ——

[a] All tuberculosis death rates in the 1930 column are for 1931, but rentals are for 1930 Tuberculosis death rates for the United States were obtained from *Vital Statistics Rates in the United States*, 1900–1940, Bureau of the Census, 1943, Table 14, p 248 Rentals are based on medians and for the United States are from *Statistical Abstract of the United States*, 1944–1945, Table 460, p 423, and Table 989, p. 912

[b] Data on the decline in rentals for the United States are not very satisfactory The general cost of living index for rent in 1930 was 98, and for 1940 was 96 This slight decline may be merely a fluctuation in trend, it may represent a deterioration in housing due to the decline in building, or it may represent a deflationary trend

($48.41) in excess of the highest rent in the low rent areas ($36 65); thus there was no overlap between the ranges of mean rentals of the experimental and control groups. Evidently a very high and consistently maintained high level of rentals must exist to be associated with a significant decline in tuberculosis death rates; low rentals, despite their decline, are not associated with a significant decline in tuberculosis death rates. Finally, the stratified random sample

(row 3) shows that a decline in rentals (from $53 86 to $37 54) is not accompanied by a significant decline in tuberculosis rates.[44]

The assumed cause-and-effect relationship between high rentals and the decline in tuberculosis death rate is measured in terms of the statistical significance of the *change* from 1930 to 1940 in rates But what *differences* between rates occur for each date singly? Table 26 (taken from rows 2 and 4 of Table 25) shows the statistical

TABLE 26 Differences between Tuberculosis Death Rates and Rentals, Experimental and Control Groups, 1930–1940

Group	1930		1940	
	TB Rates	Rentals	TB Rates	Rentals
	(1)	(2)	(3)	(4)
1 Experimental or high TB rate group	96 96	$36 65	83 87	$28 80
2 Control or low TB rate group	32 00	$72 48	22 63	$48 41
Differences	64 96	$35 83	61 24	$19 61
Critical ratios of differences	13 6	6 5	13 3	5 9

significance of differences between tuberculosis rates and rentals of each group for a given date, 1930 or 1940 It will be observed that for tuberculosis rates the critical ratio of 13 3 obtained for 1940 remained constant in 1930, when it was 13 6; similarly the critical ratios of the differences between rentals remained practically constant These evidences show that throughout the experiment the differences between the groups compared were of the same order. When the same comparison is made between the upper and the lower quartiles, the differences remain highly significant but not constant in terms of the critical ratios, a variation which is probably the result of matching.

[44] The decline in the tuberculosis death rate in the stratified sample paralleled the decline in the tuberculosis rate for the United States as a whole

New Variations in Technique Introduced in This Study

Two applications of statistical method introduced in this analysis may be noted. (1) selection of a stratified random sample from the universe to serve as a model or norm of probability for changes in rates on the independent and the dependent variables; and (2) use of partial correlation on this sample to discover the probable efficacy of specific control factors for subsequent matching.

In any experimental design the criterion of the significance of *differences between* or *changes in* rates is an estimate of the probable frequency of occurrence of such a change by chance fluctuations of random sampling. The argument runs thus: When the magnitude of the obtained difference or change in rates is one that could occur once in 100 trials (or even less frequently) in random sampling, then the obtained difference or change may be regarded as probably significant of a real difference. Now the practical difficulty with this argument is that the differences or changes in rates obtained are derived from extreme groups matched on control factors and hence not random samples. All that can be said, therefore, is that the changes in rates obtained in this study are of no greater magnitude than might occur had two random samples been chosen and a comparison made between them rather than between an experimental group and a control group. This limitation may be offset in part by selecting a stratified random sample and computing the statistical significance of changes in this sample between the two dates. Since one stratified random sample is for all practical purposes as likely to represent the universe as are two or more random samples, the stratified random sample may serve as a rough norm for comparison. Table 25 (row 3) shows that the stratified random sample does indeed serve as such a norm of probability for measures of statistical significance in approximation studies of the present sort; it serves as such a norm by showing that a change in tuberculosis death rates of 5.06 per 100,000 (from 55.30 in 1930 to 50.24 in 1940) can occur frequently in sampling (critical ratio of 0.59), but that a change of $16.32 in rentals (from $53.86 in 1930 to $37.54 in 1940) occurs very infrequently in sampling (critical ratio of 5.14);

consequently the 13 09 change in tuberculosis rates (from 96 96 in 1930 to 83 87 in 1940) in the experimental group is seen to be non-significant as compared with absolute changes in the corresponding rates of the stratified random sample, and for the control group of high rent areas the decline of 9 37 in tuberculosis rates (from 32 0 in 1930 to 22.63 in 1940) is highly significant as compared with the absolute changes registered in the stratified random sample [45]

In using the norm of probability, the definition of probability as the limit of relative frequency, or the measure of relative frequency, is used in this section. Such a definition follows the practice of Carnap, Peirce, Venn, and Von Mises,[46] rather than the definitions of meaning of Laplace[47] or of Keynes,[48] which have been sharply distinguished from the former by Carnap and Nagel

Another consideration now deserves attention Are the factors chosen for control by matching really functional factors? Here the use of partial correlation on the stratified random sample supplies an empirical answer If we hold constant one or more accessible factors and obtain an increase in the magnitude of the relationship assumed in the original working hypothesis, namely, that rentals and tuberculosis death rates are related, then symbolic manipulative control by sorting or matching on these factors is likely to be worth while This was indeed the case since $r_{15} = -.308$ increased to $r_{15\ 4} = -.66$ On the other hand, when the increase is slight, as in a similar study of the relationship between rentals and infant mortality in New York City health areas, as from $r_{12} = -.24$ to $r_{12\ 34} = -.33$ (see footnote 42), then symbolic manipulative control of these factors is less likely to isolate the real relationship. In this

[45] The underlying assumption regarding the universe sampled is that it remained homogeneous in all essential factors from 1930 to 1940 for all samples

[46] For a helpful discussion see Rudolf Carnap, "The Two Concepts of Probability," *Philosophy and Phenomenological Research*, June, 1935, pp 513–532, Ernst Nagel, "Principles of the Theory of Probability," *International Encyclopedia of Unified Sciences*, vol 1, No 6, 1939 (who distinguishes three different meanings of probability—Laplace, Keynes, and Von Mises), and T C McCormick, "Note on the Validity of Mathematical Probability in Sociological Research," *American Sociological Review*, October, 1945, pp 626–631

[47] and [48] See especially discussions by Carnap and Nagel, *op cit*, also our discussion, chap vii, pp 183–186.

manner the use of partial correlation technique to analyze relationships for the date chosen to show the effect serves as a useful and dependable pre-test of the control factors to use, and saves time and energy that might be expended in trial-and-error attempts to select good controls.

Limitations of the Present Study

Certain qualifications to the present study have been mentioned in passing, but other limitations also require attention. These may be enumerated as follows

1 In all samples only *the same health areas* of 1940 and 1930 were compared and not the same individuals or families. Only average tuberculosis rates and averages of median rentals for health areas were available and used. This limitation was not present in an earlier study, where the technique of identical individual matching was used to good effect.[49] Matching by statistical devices is effective and yet identical individual matching has the advantage of being simple to comprehend; furthermore, manual manipulation of data in this manner, although more arduous than statistical manipulation, leads to knowing the data more realistically because it is handled.

2. The present study neglects such important variables as (a) out-migration and (b) in-migration between health areas, since we found no information on these points This consideration relates to the difficult problem of selective factors discussed below.

3. Only gross rates were computed since data for specific rates by sex, occupation, and other factors were not available.

4. No specific quantitative data of a conclusive type were available on the comparative efficiency of preventive tuberculosis programs in the different health areas All that is known is that health areas of the experimental group were located in the more crowded central boroughs of the city (Manhattan and Brooklyn), whereas the health areas of the control group were located in outlying boroughs of Bronx and Queens For nine of the experimental group areas the percentage receiving out of town care, percentage cared for at home, and percentage under care of private physician (presumably those

[49] See pp 102, 118–119, 123

receiving superior medical care), were slightly lower than in 16 areas of the control group This is slender evidence of differential medical care in tuberculosis preventive work

5 The date periods differ in two cases: for tuberculosis death rates the 1931 figure (used for 1930) is the mean of 1931–1935, and the figure used for 1940 is the mean of 1936–1940. This has the effect of understating the real 1930 rate and of overstating the real 1940 rate, and thus reducing the figure of change from 1930 to 1940.

6 Perhaps the most serious limitation is neglect of the selective factor, which may have operated as follows high rent areas draw families and individuals of relatively high income, whose economic achievement *may be due* initially to greater physical vigor, stronger physical constitutions, and superior resistance to tuberculosis infection. Low rent areas receive families and individuals whose economic status *may be due* initially to lower physical vigor, weak physical constitutions, and high susceptibility to tuberculosis. This explanation is very different from proving that when a group of families or individuals is moved out of a low rent area into residence in a high rent area, other factors being equal, the tuberculosis death rate of such a moved group declines significantly *following* this move Our data do not, of course, prove or disprove this proposition. Thus the high rent areas may not operate per se to cause a significant decline in the tuberculosis death rate. The long-time decline in the tuberculosis death rate in the United States since 1900 is no doubt in part a phenomenon of natural selection, the elimination by death of susceptible strains in the population, and the survival of immune strains to form an increasing proportion of the surviving population year after year, until some stable relationship is established for any given stage of development of the arts and sciences. Such an interpretation, however, does not deprecate the valuable contributions made by preventive programs of public health and personal hygiene; nor should it be taken as an argument to weaken the support given to such preventive programs.

As we have seen in Chapter IV, the null hypothesis offers some advantages in the scientific study of the results of applying a social program because it minimizes the bias that can so easily arise in the

use of normative terms. The null hypothesis may also be used in experimental studies which attempt to test the effect of impersonal social forces like rentals in the tuberculosis study Formulated for test by the evidence of this study the null hypotheses would be:

1. There are no changes in tuberculosis death rates and rentals in New York City health areas from 1930 to 1940 when differences in the composition of an experimental and a control group are held constant on two matching factors. (See pages 130–131.)
2. The observed changes in tuberculosis death rates and rentals are not greater than those that could occur between two groups selected at random as samples from the same population.
3. The differences in observed changes in tuberculosis death rates and rentals between the experimental group and the control group are not greater than those that could occur between two groups selected at random as samples from the same population.

As we have seen in the discussion of Tables 24, 25, and 26, null hypothesis 1 is disproved. Null hypothesis 2 is disproved for changes in rental for all samples or groups, but supported by the evidence on tuberculosis death rates in the upper quartile group, the experimental group, and the stratified random sample—see Table 25, column (5) Null hypothesis 3 is disproved for rentals, but not for tuberculosis death rates

Summary Comments

The purpose of observation under conditions of control, as exemplified in the three ex post facto experimental studies just summarized, is to isolate the relationship between one independent variable, assumed to be the chief causal factor or pattern of factors, and a dependent variable, taken as the effect The operations are simplified by control through matching of several other independent variables known by prior study to be related to the dependent or effect variable.

An appraisal of the comparative success of two of the foregoing experimental tests of the results of a program, those of public housing and public high school education, seems to indicate that a rather

simple method of study, like the New Haven experiment, can yield results decisive enough to warrant repetition of the experiment

The St Paul study of the effects of high school education on economic adjustment in the community is much more elaborate than the New Haven investigation. For this reason it was possible to analyze the degree of precision with which the several control factors operated singly and in combination. Again, the experimental test of the results of the program was sufficiently decisive to encourage repetition of this form of experimental test.

But the New York City study of rentals and tuberculosis death rates is not comparable to the other two studies despite the fact that in it also the experimental design method was used. As noted on page 136, this study was severely limited by the nature of the units used—health areas and rentals It may be regarded, therefore, merely as an exploratory investigation to determine the possibility of an application of ex post facto experimental design to the testing of hypotheses about human relations in the ecological frame of reference. Yet even this study yielded sufficiently definite results, in our opinion, to suggest that it may be possible to isolate complexes of cause-and-effect relationships on a physical area basis. The investigations in human ecology have already stimulated an improvement in official recording of many social variables by small physical space areas (census tracts) in American cities. Consequently this study may open up a somewhat novel lead to further research in human ecology This hope is perhaps the chief justification for including it along with the other two more effective demonstrations of the utility of ex post facto experimental design.

Mention has been made of the need to repeat, under like conditions and on similar subjects, the kind of design herein described In Chapter X there appears a replication of the sort of study described on pages 128 to 138 of this chapter In the case of this single replicated study, similar results were obtained

Recent applications of ex post facto design are also presented in Chapter X, followed by a discussion of the sampling difficulties of the design, also, the possible application of the method of stochastic difference equations is considered

Chapter VI

SOCIOMETRIC SCALES AVAILABLE FOR CONTROL AND THE MEASUREMENT OF EFFECTS[1]

ALL shrewd observers of human relations, whether as businessmen, nurses, physicians, teachers, or trained social scientists, know that social cause and effect is extremely complex. We live in a veritable web of circumstance Factors operate functionally so that it is often difficult to decide which is the cause and which the effect. Combinations of many factors in a given social situation still further obscure the real cause-and-effect relationships But as we have attempted to show in the preceding chapters of this book, some studies have been made by use of experimental designs which open the way at least on the road to systematic and objective observation of human relations, and achieve this result to some degree by control of several of the factors operating in any given social situation. There are, of course, many unknowns still operating and still to be discovered and defined, and only time and continuous research can remedy this limitation Meanwhile it may be helpful to review briefly a list of instruments of social observation available for use in studies by experimental design

There are, in general, three groups of variable factors that need to be differentiated in the effort to unravel and observe separately the numerous factors operating in any given social situation. first, there is the dependent variable, the effect we wish to observe and describe; second, there is the independent variable that we have designated as the causal factor, and sometimes this independent variable is itself composed of several interwoven factors operating more or less as a unit pattern; and third, there are other inde-

[1] Part of this chapter was originally published under the title, "Trends in Sociometrics and Critique," by F Stuart Chapin, in *Sociometry*, July, 1940, pp 245–262, and in the author's article, "Algunos Metódos Investigación Sociológica en Los Estados Unidos," in *Revista Mexicana De Sociologia*, Enero-Avril, 1944, pp 19–36

pendent variables which, if allowed to operate freely, tend to obscure the real relationship between the first two, the relationship we wish to measure by use of experimental design of study Some of the variables in this third group are known because of previous studies of the same problem, others are unknown.[2]

All three types of factors have to be observed and measured if we hope to explain the relationships, such as housing and juvenile delinquency, counseling and adjustment, or others, which are the problems to be studied.

To be able to observe scientifically, that is, to describe and record objectively what is before us, a large number of instruments of observation are needed. For some variable factors there exist conventional records of the kind or the degree in which different persons possess these factors Such, for example, are the categories of age, sex, national language, some racial traits, income classes, occupational groups, years of school education, etc. But for many important factors there are no conventional records of kind or degree that can be used as a basis of description. These factors are often psychological, such as morale, personal adjustment, and social distance; they are sometimes obscure community influences which operate by way of gossip, habits of borrowing and lending, or as numerous folkways; and some of them are merely accepted forms of behavior which nobody has thought it important to take down or to keep a record of—and yet some of these, like memberships, attendance, committee assignments, and official positions in the small free groups and organizations of the community, are very reliable and significant indicia of leadership and other highly dynamic social forces Sociometric scales have been invented in very substantial number, as presently will appear, to permit quantitative description of these factors in human relations

The remainder of this chapter will be devoted to a description and grouping of these scales of social observation. In this connection no claim is made for the completeness of the list, and readers will

[2] Control of these unknown factors by randomization is difficult to accomplish in the free community situation, as has been noted in earlier chapters of this book, and as will be considered in chap VII, possible solutions are given in chap VIII

doubtless think of excellent scales unfortunately omitted. Our purpose is merely to show the range of observational tools already available for use by the social observer who desires to record his description of the social situation before him in an objective manner Objective records are of utmost importance if we are to avoid the insidious influence of bias and the desire to find that which we wish. Observation of the qualities of human beings in their social relationships should be stated or recorded in terms that *represent* or that *stand for* these qualities, since it is far too easy to allow our observations to take the form of terms which *express* how the observer *feels* about these qualities Thus the presumably descriptive term "communist" may be applied to a person not because the record shows that he is a member of the Communist party, has paid dues for the past ten years, and has held office for two years, but because we do not like him as a person An assumedly descriptive word becomes merely an opprobrious term to express our dislike Even a more objective term such as "moron," originally used to describe a class of individuals whose scores on a standardized intelligence test were within a certain range of magnitude, may become emotionally toned, as when one says, "Oh, he is nothing but a moron after all," thus expressing disgust at this person but by no means describing him This example illustrates one of the difficulties to calm, dispassionate, and unbiased observation of human relations. We habitually respond to human relations in terms of emotion and sentiment because these relations possess values for us. It is the achievement of psychometric tests and of sociometric scales that these instruments of observation minimize such distortions and supply the student with devices that measure human relationships.

Sociometry or Social Measurement

By the term "measurement" is meant simply description that uses numerical symbols instead of verbal or ordinary word symbols For example, it is possible to say that John is taller than Bill, thus describing the stature of these two persons by ordinary words. But when we say that John is six feet two inches (6′ 2″), and Bill is five feet two inches (5′ 2″), we have a much more precise and

verifiable form of description. For many purposes the ordinary distinction in words that John is taller than Bill suffices But if we desire more precision, then we have to resort to measurement in feet and inches. Thus measurement is a form of quantitative word description as distinguished from qualitative word description. We make the transition from one to the other whenever a scale of reference is used as the common unit of description This scale of reference has no inherent rightness or wrongness and possesses no absolute attributes It has, like other social products, evolved in practice to meet a need Some measures, like the foot rule, are the pure result of use and wont, others, like the meter stick, have been somewhat more rationally derived All depend for their validity upon general acceptance, and all are, in a certain sense, arbitrary constants accepted for their convenience Since we have been writing for some years on the subject of measurement of social phenomena,[8] we shall not labor the point further, except to say that those who protest that there can be no social measurement are, like the medieval Schoolmen, arguing a lost cause, since we now have an abundance of social measurements The claim that there can be no social measurement is like the scholastic escape from reality into the realm of dialectic, as when the determination of how many teeth a cow had was decided by the number of quotations that the orators could assemble from Aristotle upon this profound subject, rather than by looking into a cow's mouth and counting the teeth. We have looked into the cow's mouth, we have counted the teeth, and we have found that measurement is a good thing

What was really done when the cow's teeth were counted? Were the teeth measured? The cow's mouth was described by this operation, but the teeth were not actually measured Suppose in addition to counting the cow's teeth we had taken a pair of small calipers and found the width and height above the gum of each tooth That would have been true measurement. Why? The distinction is im-

[8] F Stuart Chapin, "The Meaning of Measurement in Sociology," *Publication of the American Sociological Society,* May, 1930, pp 83–94, and *Contemporary American Institutions,* Harper & Brothers, 1935, or W C Brown, 1946, chap 27, pp 353–357.

portant but often overlooked. It is this. Counting the cow's teeth merely enumerated the objects to be described. It all depends upon what is counted. *In enumeration we count objects* to be studied or to be described *In measurement we count the units on a scale* of reference The number of units on the scale that corresponds to each subject of observation (tooth) is the measurement of that subject (tooth) Perhaps the confusion arises because population count is often a fact of descriptive procedure in sociology. It is sometimes said that when we count the population of some area such as a city or a county we have measured the population of that area While it is true that such a statement is descriptive of the area, it is not measurement in any rigorous sense This is a use of the term "measurement" that no psychologist or biometrist would make. It seems that the more rigorous use of the term is less likely to lead to confusion.

Having cleared the ground of some popular confusions in usage of the concept "social measurement," we may now proceed to a classification of the field of social measurements broadly considered. It is useful to mark off three areas of social measurement first, psychometrics, or psychological measurements, which includes intelligence tests of a verbal or performance character, educational tests and measurements, personality tests and measurements,[4] attitude scales and measurements,[5] and similar measurements of the reactions of individuals as individuals, second, a division of social

[4] C Schettler, "Topical Summaries of Current Literature. Personality Traits," *American Journal of Sociology*, September, 1939, pp 234–258 Supplies 308 titles

[5] D D. Droba, "Methods for Measuring Attitudes," *Psychology Bulletin*, May, 1932, pp 309–323 Supplies 125 titles More recent summaries and scales such as the following are to be noted Hugh Carter, "Recent American Studies in Attitudes Toward War· A Summary and Evaluation," *American Sociological Review*, June, 1945, pp 343–351, H W Dunham, "Topical Summaries of Current Literature Social Attitudes," *American Journal of Sociology*, November, 1940, pp 344–375, "Methods of Personality Study A Symposium," Department of Sociology, University of Chicago (mimeographed), 1946, E W Noland, "An Application of Scaling to an Industrial Problem," *American Sociological Review*, October, 1945, pp. 631–642, Marion Strauss, "The Construction of a Multidimensional Scale to Measure Attitudes toward Labor Unions" (unpublished), M A thesis, University of Minnesota, August, 1945, and Hornell Hart, "A Reliable Scale of Value-Judgments," *American Sociological Review*, August, 1945, pp. 473–480.

measurements called "demogrametrics," or measures of large units of population such as a community, a city, or a state, and third, sociometrics proper, which is the chief subject matter of this discussion. Let us now describe in more detail these three areas of social measurement, broadly considered. In so doing, however, we shall attempt merely to indicate the different types of scales that have been developed, for to attempt any complete coverage would run up a total of considerably over 1000 titles of books and articles.

Psychometrics, defined to include educational tests and measurements, personality tests and measurements, and attitude scales, overlaps the area of sociometrics in that both subjects would include the phenomena of individual behavior and individual response in social situations. Consequently we shall pass over psychometrics with the observation that the measurement of attitudes and of some personality traits will be included in our effort to deal with sociometrics.

Demogrametrics, or the measurement of the forms or results of social behavior in large units such as the community, the city, or the state, also overlaps the field of sociometrics proper, but requires somewhat more consideration. Perhaps the earliest systematic effort to measure the comparative standing of large American cities was made by William F. Ogburn[6] and published in 1917, and a more recent systematic effort to measure the quality of living in American cities is that of Edward L. Thorndike.[7] Ogburn's study rated 36 cities on 18 criteria: child labor, church membership, cost of living, death rates, fire loss, illiteracy, infant mortality rates, library books, parks, pavement, population married, public properties, pupils to teacher, school attendance, school property, teachers' salaries, spoken English, and wage rates. Combining the 18 indices, Seattle, Salt Lake City, and Denver stood at the top; and Atlanta, Charleston, S.C., and Birmingham stood at the bottom of the list. Thorndike's 1939 rating of American cities is a much more elaborate study

[6] William F. Ogburn, "A Statistical Study of American Cities," *Reed College Record*, December, 1917, 41 pp.

[7] E. L. Thorndike, *Your City*, Harcourt, Brace, 1939, and "American Cities and States, Variations and Correlations in Activities, and the Personal Qualities of Residents," *Annals of the New York Academy of Science*, December 22, 1939, pp. 213–298.

using 37 items of comparison in contrast to the 18 used by Ogburn, but the two lists include some common and similar items, although Thorndike extends his analysis to include inter-correlations of many of his items.[8] He draws some large inferences from his analyzed ratings, for example, that *P*, personal qualities of the population of cities, are more important than *I*, income, in the determination of *G*, goodness of a city for good people. Since each of these indices is based upon factual and measured data, it may be said that Thorndike has offered an operational definition of the quality "goodness of life in cities," but here, as in all operational definitions, the definition is as good, or as bad, as the criteria he uses to validate it. Obviously it becomes a question of the degree of acceptance by scientific research students of the criteria But more of this matter of operationalism later. For the moment, we merely comment upon it in passing

Another type of rating of cities is provided by the *Appraisal Form for Local Health Work*, published by the Committee on Administrative Practice of the American Public Health Association [9] Elmer [10] states that the first complete form for field use in rating public health administration was devised by Charles Value Chapin of Providence, R I, in 1923, and that in 1924 the American Child Health Association constructed a rating schedule to express the results of its survey of health work in 86 cities.

A third type of rating for communities was devised by Nat Frame [11] in 1921. This includes separate and itemized ratings on 47 traits, some of which are matters of fact and some matters of local

[8] E L Thorndike, "Variations Among Cities in Per Capita Income," *Journal of the American Statistical Association*, September, 1937, pp 471–479

[9] American Public Health Association, 50 West 50th St, New York City, 1st ed, 1938, 185 pp and "Methods of Making Sanitation Ratings of Milk Sheds," by L C Frank, A W Fuchs, and W N Dashiell, in Reprint no 1970 from *Public Health Reports*, August 12, 1938, pp 1386–1399 U S Government Printing Office Since this earlier study, the Committee on Hygiene of Housing of the American Public Health Association has completed an extensive manual which contains devices for the quantitative description of the urban environment *Standards for Healthful Housing* vol I, *Neighborhood Environment of Dwellings, 1947*

[10] M C. Elmer, *Social Research*, Prentice-Hall, Inc, 1939, p 314.

[11] "Lifting the Country," *Extension Division Circular 255*, October, 1921, College of Agriculture, West Virginia University.

opinion, and covers such main categories as. community spirit, citizenship, recreation, health, homes, schools, churches, business, farms, and a summary score to 1000 units, being the cumulative rating on 10 classes of items with a score of 100 points each Elmer[12] devised a system for rating communities in 1924 on four groups of activities. educational, religious, recreational, and economic. Possibly the first community rating plan that gave special attention to neighborhoods was that of Willis W. Clark and J. Harold Williams[13] published by the Whittier State School (California) in 1919. This scale had five sections. neatness, sanitation, and improvements, first; recreational facilities, second, institutions and establishments, third; social status of residents, fourth; and average quality of homes, fifth. Ratings on a five-point scale were described for each of the five categories So much for efforts to rate the community, which might be a town, a small city, or a neighborhood

The fourth type of demogrametric scale is that which attempts to rate a whole state. Perhaps the most exhaustive example of this effort at measurement is the series of articles in the *American Mercury Magazine,* under the title, "The Worst American State," published in 1931[14] The first of this series of studies dealt with economics, illiteracy, and newspaper circulation; the second, with medical facts, vital statistics, some religious data, and others; and the third, with the relative standing in wealth, education, health, and the public order.

Of what use are these demogrametric studies? Can the sociologist use them as tools in scientific research? The fundamental questions are: Do these scales measure social behavior or the results of behavior with a satisfactory degree of reliability; are they valid measures of that which they purport to describe by measurement; and are

[12] *Op. cit*, pp 295–304

[13] "A Guide to the Grading of Neighborhoods," publication of the Whittier State School, Bulletin 8, July, 1919

[14] *American Mercury,* September, 1931, pp. 1–16, October, 1931, pp 175–188, November, 1931, pp 355–371 For a more recent study which supplies a basis for rating American states, see G Hirschfield and C W Strow, "Our National Health Program," *American Sociological Review,* February, 1946, pp 42–52, or Research Council for Economic Security, Chicago, Ill.

they generally accepted among social scientists? An adequate answer would require widespread use of these scales to test them out in order to prove their reliability and their validity. Obviously, only time and use can supply the answer to these fundamental questions. Assuming, however, for the sake of argument, that these scales are reasonably reliable and valid instruments of social observation, we may point out certain ways in which they could be used to advantage In the first place, they could be used to compare communities in different regions, in the second place, they could be used to measure changes in the same community over a period of years, and in the third place, they could be used to measure social factors that could be equated in control group analysis.[15] By the latter we mean this· Suppose you desire to study the behavior of delinquent boys in two communities. The first thing you need to know is whether the social environment of these two communities is the same; if you had ratings on the two communities on one or more of these scales, you could decide how near alike the two communities were. Or, to state the matter differently, you could select from a list of communities rated in this way those two communities that were nearest alike; having made this selection you would have obtained two communities with a roughly constant social environment and could then proceed to study the relationship between the amount or quality of juvenile delinquency and the operation, let us say, of the juvenile courts.

This brings us to the subject of sociometrics proper. We shall examine briefly two groups of sociometric scales first, those that attempt to measure the interaction process within social groups, and second, those that attempt to measure the family group and the home environment

The measurement of interaction within the social group may be further broken down into five sub-types of rating scale. There are first those scales or procedures that attempt to measure informal friendship constellations and seem to get at the latent culture patterns of a group of people. Illustrations of this technique are afforded

[15] See chap v, pp. 40, 109–124.

in the work of Moreno,[16] Lundberg,[17] Loomis,[18] and Franz [19] In the second place, there have been rating scales and experimental studies that center upon the description and analysis of the informal play groups of children and young persons. Illustrations of this technique are afforded by the work of Parten,[20] Thomas,[21] Arrington,[22] and Newstetter and his co-workers.[23] The third type of measurement of social interaction as embodied in its effects upon social institutions is through the study of formal organizations, and designated by a scale to measure social participation. Here we have in mind our own work on social participation first outlined in 1924 [24] and more recently described by results in 1939.[25] The fourth type of effort to measure the interaction process or its results is seen in the various

[16] J L Moreno, *Who Shall Survive*, 1934

[17] G A Lundberg and Mary Steele, "Social Attraction-Patterns in a Village," *Sociometry*, January-April, 1938, pp 375–419

[18] C P Loomis and D Davidson, Jr., "Sociometrics and the Study of New Rural Communities," *Sociometry*, January, 1939, pp 56–76, and "Social Relationships and Institutions in Seven New Rural Communities," *Social Research Report*, no 18, Bureau of Agricultural Economics and U S Department of Agriculture, Farm Security Administration, January, 1940.

[19] J G Franz, "Survey of Sociometric Techniques," *Sociometry*, October, 1939, pp 76–92

[20] Mildred B Parten, "Social Participation Among Pre-School Children," *Journal of Abnormal and Social Psychology*, October-December, 1932, pp 243–269, and "Leadership among Pre-School Children," *ibid.*, January-March, 1933, pp. 430–440, and *ibid.*, July-September, 1933, pp 136–147

[21] D S Thomas, A M Loomis, and R E. Arrington, *Observational Studies of Social Behavior*, vol 1, *Social Behavior Patterns*, Institute of Human Relations, Yale University, 1933, 271 pp.

[22] *Ibid*

[23] W I Newstetter, J J Feldstein, and T M. Newcomb, *Group Adjustment, A Study in Experimental Sociology*, School of Applied Social Sciences, Western Reserve University, 1938, 154 pp

[24] F. Stuart Chapin, "Leadership and Group Activity," *Journal of Applied Sociology (Sociology and Social Research)*, January-February, 1924, pp 141–145 In this connection also the work of W A Anderson, "The Family and Individual Social Participation," *American Sociological Review*, August, 1943, pp 420 ff, and December, 1943, pp 721 ff, deserves attention also "Family Social Participation and Social Status Self-Ratings," *ibid.*, June, 1946, pp 253–257

[25] F Stuart Chapin, "Social Participation and Social Intelligence," *American Sociological Review*, April, 1939, pp 157–166

scales to measure social distance by Bogardus in 1925,[26] by Dodd in 1935,[27] by Chapin in 1934,[28] and most recently by Zeleny[29] Finally, the fifth sub-type of measurement of social interaction or of its effects is illustrated by Williams' scale for the grading of neighborhoods in 1916,[30] and in 1933 by the extremely interesting scale by Jessie Bernard[31] to measure neighborhood folkways

Social interaction has been receiving attention in the fields of industrial sociology and in occupational sociology by Goodenough,[32] Guttman,[33] Miller,[34] and Noland.[35]

Having elaborated the measurement of interaction within the social group in terms of studies of informal friendship constellations, play groups and informal groups, formal social organizations and social institutions, and social distance and neighborhood groups and institutions, we may again ask the question, as before, To what

[26] E S. Bogardus, "Measuring Social Distances," *Journal of Applied Sociology* (*Sociology and Social Research*), March-April, 1925, pp 299–308

[27] Stuart C Dodd, "A Social Distance Test in the Near East," *American Journal of Sociology*, September, 1935, pp 194–204, and "A Test of Group Preferences in the Near East," American University of Beirut, 1933, 4 pp

[28] F Stuart Chapin, "Degrees of Kinship Intimacy," *Sociology and Social Research*, November-December, 1934, pp 117–125, and reprinted as chap. 20 in *Contemporary American Institutions*, Harper & Brothers, 1935

[29] Zeleny's many studies explore the concepts of group morale (not individual morale, cf E A Rundquist and R F Sletto, *Personality in the Depression*, University of Minnesota Press, 1936) in his "Sociometry of Morale," *American Sociological Review*, December, 1939, pp 799–808, group leaders, in his "Characteristics of Group Leaders," *Sociology and Social Research*, November-December, 1939, pp 140–149, and "Objective Selection of Group Leaders," *ibid*, March-April, 1940, pp. 326–333, and status, in his "Measurement of Social Status," *American Journal of Sociology*, January, 1940, pp 576–582 In this connection the criticisms of Zeleny's work and rejoinder should be noted in *American Journal of Sociology*, March, 1940, pp 771–776

[30] Williams, *op cit*

[31] Jessie Bernard, "An Instrument for the Measurement of Neighborhood with Experimental Applications," *Southwestern Social Science Quarterly*, September, 1937

[32] W H Goodenough, "A Technique for Scale Analysis," *Educational and Psychological Measurement*, Autumn, 1944, pp 179–190

[33] L Guttman, "A Basis for Scaling Qualitative Data," *American Sociological Review*, April, 1944, pp 139–150

[34] D C Miller, "The Social Factors of the Work Situation," *American Sociological Review*, June, 1946, pp 253–257.

[35] Noland, *op cit*, and footnote 5.

uses may these new tools be put by the research sociologist? How can they be used to solve social problems? Queen and Thomas[36] make use of the Jessie Bernard *Neighborhood Scale* in studies of the city and of urban ecology. In studies of slum clearance in Minneapolis in 1935–1936, the present author[37] used the Rundquist-Sletto scales that measure *morale* (individual morale) and *general adjustment*,[38] *social participation* and *social status*, to secure a pattern of social and individual adjustment by slum families, first in the slum neighborhood, and then after eight to ten months' relocation in different residences during slum demolition. Loomis[39] has shown how the Moreno technique of measuring friendship constellations may be used to select families for colonization in Resettlement Administration Projects of the Farm Security Administration Many of the scales cited have been used in the study of race prejudice[40]

The second group of sociometric scales to be considered are those that measure the family group and the home as a social environment Two sub-groupings of these scales may be conveniently made: The first deals with attitudes and interests of members of a family, and the work of Clifford Kirkpatrick[41] may be cited as an illustration of this approach; the second includes some 19 scales devised to measure various aspects of status, defined as socio-economic status or as social status Since the work of Kirkpatrick in measuring the

[36] S A Queen and L F Thomas, *The City*, McGraw-Hill, 1939, pp 100, 308, 370 See also measurement of community satisfactions in E C McVoy and Lowry Nelson, *Agricultural Extension Bulletin 370*, University of Minnesota, 1943

[37] F Stuart Chapin, "The Effects of Slum Clearance and Rehousing on Family and Community Relationships in Minneapolis," *American Journal of Sociology*, March, 1938, pp 744–763

[38] These scales were published in E A Rundquist and R F Sletto, *Personality in the Depression*, University of Minnesota Press, 1936

[39] *Op cit*

[40] D Young, *American Minority Peoples*, Harper & Brothers, 1932

[41] Clifford Kirkpatrick, "The Measurement of Ethical Inconsistency in Marriage," *International Journal of Ethics*, July, 1936, pp 444–460, "Community of Interest and the Measurement of Marriage Adjustment," *The Family*, June, 1937, and "Content of a Scale for Measuring Attitudes Towards Feminism," *Sociology and Social Research*, July-August, 1936, pp 512–526, and by his students in M A theses as follows B Rosenthal, "Standardization of a Community of Interest Scale, as a Measure of Marital Adjustment," June, 1939; and L Gross, "Construction of a Belief Pattern Scale for Measuring Attitudes toward Romanticism," August, 1939

attitudinal complex of family life falls in the overlap zone of psychometrics and sociometrics, we merely mention his work in passing, and turn to a more detailed discussion of the various types of scales that measure socio-economic status, broadly considered.

Scales that attempt to measure socio-economic status rate on one feature (occupational class)[42] or upon several features,[43] and may require a visit to the home to procure the information,[44] or may be secured without a home visit [45] Furthermore, in all of these scales, the economic or material culture features weigh heavily. Sometimes there are items that include inquiries about occupation, neighborhood (McCormick, Leahy), often there are items that include education, school attendance, books, magazines (McCormick, Leahy, Sims); and sometimes only one of these items is included, as books or magazines (Chapin, Morgan-Leahy). In all of them weights are assigned to the presence of items deemed to be indicative of the level of status of the family, and penalties, or negative weights, are occasionally given for a few items (Chapin). Some include articles of furniture or housing conditions (Chapin, Commons) No generalization about the type of items can be made since the range of items is large On the other hand, those scales that have been standardized are commonly the product of systematic research that involves such familiar and conventional procedures as the determination of their reliability and validity, and item analysis to secure an objective basis of the item weights [46] In regard to the method of scale construction

[42] Scales by Taussig (1913), Barr (1918), Kornhauser (1918), and Anderson and Goodenough (1931)

[43] Burdick (1928), Chapman-Sims (1925), Chapin (1928, 1933, 1936), Commons (1908), Engel (1895), LePlay (1885), Leahy Shea (1935), Holley (1916), McCormick (1929), Perry (1913), Williams (1918), Sewell (1939), E L Kirkpatrick (1924, 1933), Sydenstricker and King (1917), and Zimmerman (1924)

[44] Scales of Chapin, Commons, Leahy, McCormick, Perry, Williams, Sewell, E L. Kirkpatrick, Sydenstricker and King, and Zimmerman

[45] Scales of Barr, Burdick, Chapman-Sims, Holley, Kornhauser, Sims, and Taussig do not require a home visit, but may be filled out on interview or in school by any informed member of the family

[46] Consult Mary J McCormick, *A Scale for Measuring Social Adequacy,* October 15, 1930, *Social Science Monographs,* vol 1, no 3, National Catholic School of Social Service, Washington, D C , Alice M Leahy, *The Measurement of Urban Home Environment,* University of Minnesota Press, 1936, V. M Sims, *The Measurement*

used, we can generalize to the extent of saying that those scales concerning which coefficients of reliability and coefficients of validity have been published, based upon good samples, are tools of observation that can be used without hesitation in social research Usually also, the norms of measurements are published By norms we mean the average scores on a given scale usually found for different age groups, occupational groups, income groups, relief groups, nonrelief groups, or slum groups. By reference to these norms the research worker can determine the departures, if any, of the group of subjects he is investigating, from the expected

To what practical or theoretical uses have these scales been put? They have been used to compare families in slum areas with families in middle-class areas, to compare foster homes with natural homes, to compare one neighborhood of a community with another, and to measure the change in status during depression years.[47] They may be used in experimental studies in the community to control the factor of social status in comparisons before and after rehousing, before and after moving, etc.[48] Since they have been used to measure social status before and after some social action program has been put into operation, they may to this extent be used to measure social adjustment to changed conditions When several sociometric scales are used simultaneously, in "before" and "after" measurements, they may be used (with limitations) to measure a pattern of social adjustment [49] From a theoretical point of view, these scales may be used to

---

of Socio-Economic Status, Public School Publishing Co, Bloomington, Ill, 1928, F Stuart Chapin, *Contemporary American Institutions*, Harper & Brothers, 1935, chap 19, being a reprint of a pamphlet of 1933, and W H Sewell, "The Construction and Standardization of a Scale for the Measurement of Farm Family Socio-Economic Status, University of Minnesota Ph D thesis, 1939 The most recent standardized scale to measure urban housing is *An Appraisal Method for Measuring the Quality of Housing* (a yardstick for health officers, housing officials, and planners), Part I, "Nature and Uses of the Method," Committee on Hygiene of Housing, American Public Health Association, 71 pp, 1945 For rural housing, see Howard Cottam, *Housing and Attitudes toward Housing in Rural Pennsylvania*, Pennsylvania State College, Bulletin 436, 1942, Agricultural Experiment Station, C I Mosier, *Evaluating Rural Housing*, University of Florida Press, Gainesville, Fla

[47] See chap IV, footnotes 9, 11, 13.

[48] See chap. IV, pp 58–79.

[49] See chap III, pp 41–50

obtain an operational definition of the concept "social status," previously defined only in verbal terms. The same generalization may be made of all other sociometric scales properly standardized. On the whole, the more carefully calibrated scales, that is, those that enable the student to measure finer gradations of any measured trait among a group of subjects, are the attitude scales and the social status scales [50] The demogrametric scales are still rather crude instruments. In general, one may judge the dependability of any scale by ascertaining whether the standardization procedures, item analysis, reliability coefficients, validity coefficients, and norms have been published.[51] Reliability coefficients should usually be at least +.80 and preferably over +.90, and should be based upon correlations of first and second ratings on the same subjects by the same or by different raters, rather than on split-half correlation. Validity coefficients are a much more difficult and technical problem, as Bowers[52] and Chapin[53] have shown Validity coefficients are usually smaller than reliability coefficients, and correlations ranging from +.40 to +.79 are usually dependable. Validity coefficients as high as +.90 are seldom desirable because they can signify that since the criterion scale (validating scale) measures the trait as well as the new scale does, it may have been unnecessary to construct the new scale at all, since the trait has already been measured satisfactorily by an existing scale.

The next steps in the use or application of sociometric tools are: (1) in the operational definition of sociological concepts and (2) as instruments to provide measured controls, or as criteria of effect, in experimental studies.

[50] Descriptions of the social status scales not identified by footnote citation in this article will be found in the monographs cited by Chapin, Leahy, McCormick, and Sims

[51] For a discussion of some of the trait scales see F Stuart Chapin, "New Trends in Social Research—Some Hypotheses and Some Sociometric Scales," *Journal of Educational Sociology*, May, 1938, pp 561–568

[52] R V Bowers, "An Analysis of the Problem of Validity," *American Sociological Review*, February, 1936, pp 69–74, based upon R F Sletto, "A Critical Study of the Criterion of Internal Consistency in Personality Scale Construction," *ibid*, pp. 61–68

[53] F Stuart Chapin, "Definition of Definitions of Concepts," *Social Forces*, December, 1939, pp 153–160.

Recently there has been considerable discussion about the possibility of applying to the clarification of sociological concepts the operational procedures found so effective in physics. Since there is not time in this brief section to discuss this problem and the growing controversy that surrounds it with any degree of thoroughness, we shall merely call attention in this connection to some work by Alpert,[54] Blumer,[55] Chapin,[56] Dodd,[57] Lundberg,[58] Merton,[59] and Stevens[60] Our own point of view can be stated briefly by quoting from a later article, "Definition of Definitions of Concepts," published in the December, 1939, issue of *Social Forces*. The whole problem clarifies itself when it is recognized that the so-called operational definition is not set up as any final or absolute solution, but merely as a useful development in the direction of objectivity, and needs to be oriented against the background of sociological literature that will always have to rely heavily upon qualitative word definitions.

The problem is this since the operational definition of a social concept depends upon the series of acts performed by the investigator in the process of measurement, how can one be sure whether the scale of measurement used really does measure the subject or the objects to which it is applied? This is the problem of validity of the scale of measurement. But the operational definition that science holds up as a desired goal of objectivity states that the measurement is the concept.

Let us consider an illustration of this point In the trend toward quan-

[54] H Alpert, "Operational Definitions in Sociology," *American Sociological Review*, December, 1938, pp 855–861

[55] H Blumer, "The Problem of the Concept in Social Psychology," *American Journal of Sociology*, March, 1940, pp 707–719.

[56] F Stuart Chapin, "Definition of Definitions of Concepts," *Social Forces*, December, 1939, pp 153–160

[57] Stuart C Dodd, "A System of Operationally Defined Concepts for Sociology," *American Sociological Review*, October, 1939, pp 619–634

[58] G A Lundberg, "The Thoughtways of Contemporary Sociology," *American Sociological Review*, October, 1936, pp 703–723, and "The Measurement of Socio-economic Status," *ibid*, February, 1940, pp 29–39, and *Foundations of Sociology*, Macmillan, 1940

[59] R K Merton, "Fact and Factitiousness in Ethnic Opinionnaires," *American Sociological Review*, February, 1940, pp 13–28

[60] S S Stevens, "The Operational Basis of Psychology," *American Journal of Psychology*, April, 1935, pp 323–330, and "The Operational Definition of Psychological Concepts," *Psychology Review*, November, 1935, pp 517–527

titative description or measurement in social psychology and sociology, we strive to define each concept in terms of the measurements upon it. Scales to measure public opinion are constructed and tested for validity. We say of a scale that has been standardized to measure public opinion, "Public opinion is what this scale measures" This is then the operational definition of the concept public opinion Does this statement of operationalism make meaningless or false the basic question of validity, which is "Does this scale measure that which it is designed to measure, namely, public opinion?" This apparent dilemma appears at first glance to be destructive of the very basis of operationalism, because the proposition appears to beg the question by assuming the conclusion which is to be proved and making it part of the premises used to prove it In reality, however, the dilemma is not a real one, because the assertion, "Public opinion is what this scale measures," is made only *after* the scale has been standardized The process of standardization, if done thoroughly, disposes of the question of validity, so that the assertion of the operational form of the definition of public opinion does not beg the question.

We come finally to the use of sociometric tools as a means of providing measured controls or objective criteria of effect in experimental studies

In the use of sociometric tools to obtain measured controls in studies of matched groups, it may be pointed out that if one desires to hold constant the factor of social status in a comparison between families with delinquent boys and families not having delinquent boys in order to eliminate the influence of status on behavior, the procedure is to select for comparison those families with the same scores on social status as measured by one or more of the scales noted in this section. This simple illustration of the principle may be extended to the use of a battery of sociometric scores to control or limit the variability in certain factors so as to be able to observe the relationship between two other free variables whose cause and effect or other relationship we wish to isolate. This procedure is often used in the form of matching an experimental group and a control group on the occupational factor, in order to observe the relationship between two other variables studied, with the occupational factor held constant.

Summary Classification of Sociometric Scales

1. Psychometric Scales psychological measurements including
   (1) intelligence tests
   (2) personality tests and scales
   (3) attitude tests and scales
2. Demogrametrics measurement of the forms or results of social behavior in large units such as the community, the city, and the state
   (1) the quality of living in American cities (Ogburn, Thorndike)
   (2) appraisal of local public health (A P H A , and C V Chapin)
   (3) rating smaller communities (Frame, Elmer, Williams)
   (4) rating an American state (*American Mercury,* Hirschfield)
3. Sociometric Scales measurements of the social process
   (1) interaction within the social group
      a. measurement of informal friendship constellations (Moreno, Lundberg, Loomis, Franz)
      b measurement of informal play groups of children (Parten, Thomas, Loomis, Arrington, and Newstetter)
      c measurement of activity within formal groups (F S Chapin, W Anderson)
      d measurement of social distance (Bogardus, Dodd, F S Chapin, and Zeleny)
      e measurement of neighborhood influences and occupations (J. Bernard, Goodenough, Guttman, Miller, and Noland)
   (2) attitudes and interests of members of the family (Kirkpatrick, Gross, etc)
   (3) social status, socio-economic status, and home environment (older scales by Commons, Engel, LePlay, Taussig, Barr, Kornhauser, Chapman-Sims, Burdick, and newer scales by Anderson and Goodenough, McCormick, Chapin, Sewell, etc)

AN ILLUSTRATION OF AN APPLICATION OF SEMANTICS AND SYNTACTICS TO THE IMPROVEMENT OF SOCIOMETRIC TOOLS[61]

Before concluding this discussion of sociometric scales as tools or instruments of social observation, brief consideration may be given

[61] This section appeared originally as an article, "The Syntactical Analysis of Sociometric Techniques Cases in Point," by F Stuart Chapin in *Sociometry,* May, 1941, pp 177–183

to the use of a new language tool, semantics and syntactics, for the criticism of the items in sociometric scales Since the social scientist obtains instruments of observation (tools of observation) by constructing sociometric scales, and many of these scales consist of linguistic items rather than of material substances (like the instruments of precise observation of physical science), the student who constructs a new scale must utilize language items with great care. Furthermore, since much social behavior consists of language behavior, vocalization, writing, and linguistic habits of response, complications easily occur when the language behavior of the student (as social scientist) is used to describe and record the language behavior of the subject, the person behaving. Thus we have before us the problem of how to use the tool of language to create a tool of observation to be used in the study of linguistic behavior. In semantics and syntactics we have a key to this problem.

Syntactical analysis is a powerful tool for the dissection of theoretical statements and working hypotheses of social research This section will examine two cases first, an assertion about attitude measurement that represents a fairly common philosophic point of view; second, one aspect of the controversy over units of measurement. It is hoped that this brief analysis will clarify some aspects of method. Clarification of method is useful when it furnishes reliable and valid tools or procedures that many other research workers can use with confidence Such a discussion is not an escape from reality into the petty refinements of technique when specific facts are used to test the points clarified and when those who read the results are left to make their own inferences as to validity and utility.

A Case in the Measurement of Attitude

A statement in the authentic language of social science by a well-known author about attitude measurement is, "They seek to apply *mechanical methods* of measurement to things whose *very nature* they fail to understand" [62] (italics ours).

Analysis of this statement shows that two words, "mechanical methods," function in this sentence with invidious purpose. They

[62] Robert M MacIver, *Society Its Structure and Changes*, 1931, p. 45.

are a familiar stereotype used in terms of reproof. Second, the statement also uses the other words, "very nature," to imply a metaphysical presupposition. Stereotypes and metaphysical terms are permissible in ordinary language behavior (the folk language) as *subject matter of study,* but are not permissible in the language of science used as *a tool* to describe or to analyze ordinary language behavior. Hence this sentence is what Carnap would call a sentence in the pseudo-thing language and is non-sense. By contrast, the thing sentence that this MacIver statement could be *reduced* to in order to make it have sense, in the empirical science meaning, would be: "They seek to explain attitudes by measurements on scales of such grossness of calibration that the instrument misses differences that may be highly significant."

Probably MacIver would agree that this was what he really meant. He did not, however, state his criticism in this way, but used the all-too-familiar pattern of philosophic language in a context in which an empirical science statement would be expected. We make this last assertion because the original statement is not capable of proof or disproof, whereas the statement in its reduced form is capable of proof or disproof. If the scale in question differentiates between individuals, the proposition is disproved That such proof can be demonstrated is shown in the analysis of item 47 from the Rundquist-Sletto [68] scale of opinions where relief clients have a 44-per-cent more radical scale position than employed persons

In item 47 note the use of the word "ought" (italics ours) This word does not make any further statement or assertion about the matter, but is merely equivalent to saying, "The government guarantees a living wage to those who can't work" in a tone of approval or with greater emphasis. The word "ought" is thus merely a word that functions in the sentence to express emotion The use of this word makes the statement a normative statement incapable of proof, and therefore a pseudo-thing sentence. The word "ought" serves as a normative or ethical symbol, the function of which is emotive, to express emotion or to elicit an emotional response in the subject

[68] E A Rundquist and R F Sletto, *Personality in the Depression,* pp 384–391, and 394–398.

(the person who responds to the test); as such it is a good statement in the folk language because it is a stimulus to which different degrees of emotional response can be recorded by setting the symbols: 5 = strongly agree, 4 = agree, 3 = undecided, 2 = disagree, and 1 = strongly disagree. Thus the number symbols stand for or represent different degrees of emotion elicited by the stimulus, and these numerical symbols are susceptible of arithmetic and statistical manipulation to obtain averages for a group.[64]

Sentences from the folk language	"47 The government *ought* to guarantee a living to those who can't find work."		"29 Poverty is a result of *injustice* in the distribution of wealth."
The language of science used as a tool to describe behavior in the folk language	5 4 3 2 1		5 4 3 2 1
	Mean scores	(1)	Mean scores
	2.72	Employed	2.89
		(2)	
	3.24	Unemployed	3.14
		(3)	
	3.94	Relief	3.54
	1.22	Differences	0.65

Item 29, in which the stereotype "injustice" (italics ours) serves in an emotive function, is not as effective in differentiating between employed and relief groups as is item 47. Nevertheless both items have critical ratios of over 2, and number 47 is statistically significant. To return to our former point, we may now assert that we have proved that item 47 does not miss differences that are significant, and that therefore this item has a fineness of calibration that distinguishes between the attitudes of employed and relief persons, two groups that would be expected to have different attitudes on this issue. We have thus disproved the statement about attitude measurement when rendered in its *reduced* form, a demonstration

[64] This problem has been recently explored under the concept "the intensity function" in attitude research by Guttman, Goodenough, and Noland, *op. cit.*

that could never have been made in the case of the original statement in its philosophic phraseology. Hence the value of syntactical analysis applied to the dissection of a loose statement.

Equal and Interchangeable Units of a Scale

Our second case concerns the controversy over units of measurements in psychometric and sociometric scales. The proposition that these units are equal and interchangeable has been claimed and denied by Kirkpatrick, Merton, J. Bernard, and Lundberg.[65]

This controversy arises for two reasons: first, because of the absence of a clear distinction between (a) units of subject matter or objects of study, which we have designated "units of observation"; [66] and (b) units of reference which are arbitrarily assumed equal for purposes of convenience in linear continuum, which we have previously designated "units of measurement."

The controversy arises in the second place because the foregoing distinction between units of observation and units of measurement is difficult to grasp when applied to intangible social attitudes.

Let us begin with the first reason for the controversy Units of observation, (a), may be similar but can never be equal or interchangeable (one human being[67] is more like another than is any human being like an elephant). On the other hand, units of measurement, (b), on an assumed linear scale are equal and interchangeable distances or numerical weights representing these distances. These units of distance or weights *are arbitrarily set as equal* to assumed degrees of approval or disapproval in attitude measurement. The only way to justify such an assumption is by *using* a scale in which such units are a part of the continuum, and then *finding* that such

[65] Clifford Kirkpatrick, "Assumptions and Methods in Attitude Measurement," *American Sociological Review*, February, 1936, Robert K Merton, "Fact and Factitiousness in Ethnic Opinionnaires," *ibid*, February, 1940, pp 13–28, J Bernard, *ibid*, June, 1940, G A Lundberg, *ibid.*, February, 1940, pp. 38–39, and R. K Merton, *ibid*, August, 1940

[66] F Stuart Chapin, "The Meaning of Measurement in Sociology," *Publication of the American Sociological Society*, May, 1930, pp 83–94 (especially pp 89–92)

[67] It is customary to count the number of persons in two different areas and take the resulting sums as measures of population This use of the concept measurement is loose, as we have pointed out in previous analyses of the concept.

a scale does in fact differentiate by significantly different scores between two groups that are recognized as different on some other accepted device of differentiation. For example, when the Rundquist-Sletto scale shows a mean score of relief persons of 3.94 on item 47, and a mean score of 2.72 for employed persons, the difference, 1.22, is so large as to be scarcely due to pure chance (as understood by reasonable persons). Furthermore, one knows that employed persons and relief clients are in different social situations in terms of economic security Hence whatever the crudities of the original assumption of linear units, *the scale does work* as a differentiating device. As long as it works it is a useful instrument.

Since the distinction between (a) units of observation and (b) units of measurement is especially difficult to grasp when applied to such intangible things as social attitudes, the arguments pro and con are not easy to substantiate in terms sufficiently concrete to be convincing. We shall next consider, therefore, an example in which the units of observation are not abstractions like social attitudes, but tangible sense data. Take the case of measurements with our social status scale.

Louis Guttman [68] in 1942 completed a multiple factor analysis of the social status scale measurements on 67 Negro families in Minneapolis and St. Paul made in 1932, to test the validity of the numerical weights used in the 1932–1933 edition of the scale.[69] In this scale, which lists articles of furnishings in an ordinary living room, there are 17 items. To explain concretely the distinction between units of observation and units of measurement let us take as examples item 15, newspapers, to which a weight of 8 is assigned, and item 17, radio, to which a weight of 8 is also assigned. The question is, was a weight of 8 the correct weight of each such item if the scale is to

[68] For an abstract of the results of this WPA factor analysis study under the direction of Mr Guttman see footnote 15, chap v It may be stated that this analysis involved the computation of 4305 correlations over a period of more than a year and a half

[69] F. Stuart Chapin, *Contemporary American Institutions*, Harper & Brothers, 1935, W. C Brown, 1946, pp 378–382. In addition to the 17 items of sense data, there are 4 items of opinion, but these items also have satisfactory reliability and validity as tests show, and hence do not affect the argument given herein.

differentiate on total score of the living room between families of different income class, different occupational group, differences in total years of formal education of the parents, and differences in the participation scores[70] of the parents in organized groups of the community?

Guttman shows by factor analysis that new (more valid) weights in comparison with the old (arbitrary) weights may be derived. For instance

	Units	Old Weight	New Weight
4	Fireplace with 3 or more utensils	8	34 6
15	Newspapers (each)	8	7 7
16	Telephone	8	24.4
17	Radio	8	8 0

It is evident from inspection of these four items: (1) that the original weights of items 4 and 16 were quite inaccurate and far too low, and (2) that the original weights of items 15 and 17 were about correct

What is the bearing of this finding upon the issue of "equal and interchangeable units"? It is this Factor analysis reveals that the presence of a fireplace and 3 or more utensils (number 4), and the presence of a telephone (number 16), are more significant indicators of differences in social status on four other indices of social status (income, occupation, education, and participation), than we had originally assumed But factor analysis also proves that weights of 8 each for newspapers and radio were valid weights[71]

In this example it seems that the objects "newspaper" and "radio" are, when translated into numerical weights, equal and interchangeable units in two meanings: (1) as units of observation, subjects of study to which sensory reactions are made, and also (2) as units of

[70] F Stuart Chapin, "Social Participation and Social Intelligence," *American Sociological Review*, April, 1939, pp 157–166

[71] Further verification of the new Guttman weights is obtained by using them against the old weights in computing a critical ratio between an experimental group of public housing residents and a matched group of slum families for differences in gains on social status from 1939 to 1940, where the critical ratio was increased from 0 8 to 1 77

measurement, equal and interchangeable numerical weights or distances on an assumed linear scale The same statement cannot, however, be made of item 4 and item 16, since by Guttman's analysis the importance of the former is 1.4 times the importance of the latter, if we accept the description of social status on the combined evidence of income, occupation, education, and social participation as an external criterion of social status. Thus although two of the items (objects) are of about equal importance numerically, the other two items are not of equal importance numerically. This indicates that in one simple scale at least, *there is no necessary intrinsic correspondence between the weights (units of measurement) and the objects or subject matter (units of observation)*. It is purely coincidence that two of the items (newspapers and radio) should turn out to have the same numerical values in the two validations of the scale against external criteria. This analysis does, it seems to us, help clarify the basis of the controversy over the condition that units are equal and interchangeable Incidentally, it seems to provide evidence in support of our original position of 1930, when the distinction in regard to social measurements was first made, that (1) units of observation (subject phenomena) are not to be confused with (2) units of measurement (reference phenomena), if social measurement as a useful device is to progress to a more precise form of scientific description.

A number of new methods of resolving some of the problems of social measurement have reached a stage of development at which helpful adaptations could be made in experimental designs. Consequently, the reader may wish to note in particular: (1) methods of overcoming the obstacles to randomization in the free community situation in Chapter VIII, (2) methods of testing experimental designs with fewer assumptions, especially those of Friedman, Kruskal-Wallis, and Haberman in Chapter IX; (3) possible use of the method of stochastic difference equations in Chapter X; and (4) the statistical theory of extreme values, and the Guttman techniques and theory in Chapter XI.

Chapter VII

SOME FUNDAMENTAL PROBLEMS AND LIMITATIONS TO STUDY BY EXPERIMENTAL DESIGNS

IN THE preceding chapters we have noted the fact that there are many unsolved problems and some fundamental limitations to be overcome in studies by experimental design. Systematic consideration of these difficulties has been deferred for analysis in this concluding chapter These problems and limitations are of two kinds· first, practical obstacles, and second, theoretical limitations that derive from the conditions of the probability theory wherever critical ratios are used as a provisional test of the statistical significance of differences in means, changes in means, or differences in changes in means of such non-random samples as an experimental group and a control group We shall now consider the practical obstacles, and then pass to a consideration of the theoretical limitations

PRACTICAL OBSTACLES[1]

Field interviews of individuals in their homes and in the community setting reveal problems not present in experimental designs that are set up in the comparatively simple schoolroom situation,[2] familiar in educational research. The problems of field interview emerging in the community situation in which research uses some experimental design are both methodological and practical. In this section we shall attempt to describe these problems as they arose in three research studies using three different variations of experi-

[1] This section appeared originally in an article, "Some Problems in Field Interviews When Using the Control Group Technique in Studies in the Community," by F Stuart Chapin, *American Sociological Review*, February, 1940, pp 63–68

[2] E F Lindquist, *Statistical Analysis in Educational Research*, Houghton Mifflin, 1940, pp 80–84.

mental design. Wherever possible, solutions of these problems will be indicated.

To clarify our analyses it should be stated at the outset that by problems of field interview we mean problems of organization and timing of these interviews and not problems of minute personal relationship in interview situations Since some study has been made of problems that arise in face-to-face contacts of interviewing,[3] and less attention has been given to problems of organization and timing, it is with these latter problems that we shall be concerned, and more specifically with scheduling interviews in home calls chiefly as related to sampling.

The purpose of experimental designs applied to the study of social factors is to measure the effect of one social factor at a time, all other factors in the situation (unknown as well as known) being held constant, that is, controlled. Difficulties arise at once because it is seldom possible to measure one factor alone Usually the best that we can do is to try to measure a group of factors which operate as social treatment, or as a social program—as, for instance, in an effort to isolate and measure the effect of improved housing upon a group of families formerly slum residents.[4] A second difficulty arises because there are unknown factors at work, that is, unmeasured factors which are not controlled.

Randomization[5] is theoretically the solution to obtaining control of these unknown factors The procedure is to select a sample at random from the population to be studied. This sample may then be divided at random into the experimental group, which is to receive the social program (e g, good housing), and the control group, which is excluded from the program (e g, the slum dwellers). Or two random samples are selected, one to receive the program, the

[3] Twila E Neely, *A Study of Error in the Interview*, Columbia University Press, 1937; see also several later studies in *Sociometry*, May, 1942, pp 109–134 See also R K Merton and Patricia L Kendall, "The Focused Interview," *American Journal of Sociology*, May, 1946, pp 541–557

[4] See chap iv, pp 58–79

[5] Lindquist, *op cit*, pp 24–29, T. C McCormick, *Elementary Social Statistics*, McGraw-Hill, 1941, pp. 27–28, Quinn McNemar, "Sampling in Psychological Research," *Psychology Bulletin*, June, 1940, pp 331–335, 353

other to act as a control and to be excluded from the program. Since each random sample is by definition selected without bias it will probably contain the unknown or unmeasured factors in equal degree.

The next step is to measure simultaneously each group on some criteria designed to evaluate the effect of the program (e.g., scales that measure morale, adjustment, social status, social participation, etc). After an interval of a year or more, each group is again measured. Then if the difference in measured changes between the experimental group and the control group shows that the experimental group gains in a degree that occurs very infrequently in chance, the odds are in favor of concluding that we have discovered a difference that is probably a measure of the effect of the social program.[6] So runs the argument.

In community research in practice, however, we encounter an initial difficulty to this apparently smooth logical conclusion. Some cases disappear during the interval of time that the experiment runs. Some are lost by refusals to give information on second interview, some have become deceased, and some have moved away and cannot be found. As a consequence the terminal groups are not composed of the same individuals as at the outset. This means that at the terminal date the groups have lost their random character. Certain selective influences have been at work

A possible solution of this dilemma is to select several random samples for experimental groups and several for control groups, on the assumption that at least one pair of samples will not lose cases and hence will remain random samples throughout the experiment. The practical difficulty to be overcome by this remedy is that of cost. When several samples or pairs of samples are used it is necessary to interview many more cases and this procedure rapidly runs up the expense of research. It increases the cost also because a larger number of interviewers have to be trained and supervised.[7]

[6] *Disproof of null hypotheses is assumed as part of the procedure*

[7] In the housing study cited, the interviews required from 35 to 50 minutes each to fill in the three schedules or scales used, from which the four measures were derived For a sample of 120 cases, 6 interviews a day would require 20 days, and 10 interviews a day, 12 days.

More serious than this difficulty is another very practical and fundamental one. It is this. Will the directors of a social program be willing to give their treatment to a randomly selected group and exclude from treatment another identical group to provide a theoretically sound experimental design? When we were asked to study the effects upon the morale of clients resulting from a WPA program in comparison to that of general relief (local relief), it was not possible to use a procedure of this sort.[8] Who ever heard of a director of WPA or of relief or of public housing willing to court the public criticism that would develop if it became known that selection for treatment or exclusion was made on a random basis? In the WPA study a public controversy was already under way relative to the continuance or limitation of the program. The public atmosphere was therefore not conducive to strengthening support for a research procedure which seemed to favor certain individuals at the expense of others. People resent being chosen to serve as "guinea pigs." Would a government administrator permit admission to a public housing project of some families and exclusion of others equally eligible on the basis of random choice? Most administrators are charged with the responsibility to admit to good housing only the most needy families, other factors being equal. This is the way social reform programs are set up to function because those who sponsor them never question their beneficial effects any more than they ever expect to bring such programs to a scientific test of their effects. In the housing experiment cited, there had been much criticism of the project from conservative groups of the community and this created for the administrators many practical difficulties. No public administrator would like to be in the position of seeming to favor one group (in this case the experimental group to be selected at random to have the benefit of treatment) at the expense of another group (in this case the excluded control group, also to be selected at random), without tangible evidence of the greater eligibility on the part of the beneficiaries of the program. Once greater eligibility is accepted as a criterion of admission, the randomness of the group disappears,

[8] See chap. III, pp 41–50. For a suggested solution see chap. VIII.

and with it one of the essential conditions of an ideally theoretical experiment.

Until public administrators of social programs can see their way around this problem, it appears that the use of randomization as a method of control of unknown factors can be ruled out in experimental designs as a method of evaluating social programs in the community situation (For recent solutions see Chapter VIII)

The research student turns, therefore, to the control group technique This method limits the use of randomization to the selection of the control group We are obliged to accept as the experimental group any accessible portion of the population eligible to and receiving the program, whether a WPA, relief, or housing program. The control group consists of a random sample selected from a similar population not receiving the program. This was the plan followed in the WPA study cited In the housing study, for reasons of time limitation placed upon the research, it was not possible to select the control group at random In this case the control group, like the experimental group, consisted of accessible families living in the slum in the same residence for the duration of the experiment.

To reduce variation among other factors the two groups are then matched on as many known factors as possible. These may be age, sex, size of family, income, occupational class, etc. The experiment proceeds from this point as before measurements are made before and after on selected criteria of effect. Since individual cases are lost during the interval of the experiment, it is important to ask, "What are the reasons for these eliminations?"

Our experience in two studies using before and after measurements on both an experimental and a control group shows that the longer the experiment runs in a free community situation, the larger the number of cases lost. The first of these two studies was an ex post facto experiment[9] The purpose of this study was to measure the effect on individual adjustment to community life of the factor, "length of high school education," on the hypothesis that the longer the term of high school education the better the adjustment to com-

[9] See chap v, pp 99–124

munity life nine years after graduation or leaving school We shall not repeat here a full account of this research, since it has been described in Chapter V. The study began with the school records of 2127 St. Paul high school students of 1926. Of these, 1130 had graduated in 1926 after a four-year course, and 997 had in 1926 dropped out of school, in which year they had completed the first, or second, or third year By 1935, or nine years later, an interview follow-up in the community situation of these 2127 young people found only 1194 cases During the nine-year interval between leaving school in 1926 and the year 1935, when field interviews began, some 933 of the original 2127, or 43 8 per cent, were lost for further study. The reasons for these losses were: 575 could not be found, 42 had moved out of the community, 21 were deceased, and for 295 the records were so incomplete as to prevent using these cases in further comparative study. Of the experimental group of 1130 graduates in 1926, some 459, or 40 6 per cent, were among those lost for the reasons stated above Of the control group of 997 school drops or incomplete, 474, or 47 5 per cent, were lost. In this instance, as in other similar studies, the losses were greater from the control group than from the experimental group.

In the second study, the housing study mentioned above, we began with measures of morale, general adjustment, social status, and social participation in 1939, on 108 families admitted to Sumner Field Homes of Minneapolis, and a control group of 131 similar families living in the slum. A total of 48 families were lost before 1940 when the experimental period ended. Of this number, 7 dropped out of the experimental group because they could not be located and 41 were eliminated from the control group The breakdown of these 41 dropped cases was 30 not located after careful field work follow-up, 5 moved out of town, and 6 moved into the housing project (hence had to be excluded from the control group). In all, 20 per cent of the initial number of families were lost during the experimental period of one year.

But analysis of the drops due to mobility and miscellaneous reasons does not tell the whole story. In the housing study additional cases were lost owing to refusals to give information after repeated

and courteous attempts at interviews. A total of 20 cases were lost for this reason 6 from the experimental group of residents and 14 from the control group of slum families. Additional miscellaneous reasons accounted for a loss of 7 cases from the control group. In summary, a total of 75 cases were lost[10] in these ways from the 239 original total of both groups. (See Chapter IV, pp 58–79) This suggests the degree to which samples would have been distorted from representativeness had it been possible to select the individuals at random. The original samples were not chosen randomly because, as explained above, a social program conducted in the milieu of the contemporary community folkways and mores cannot be applied to random groups of the population. The regulations that express the purpose of such a program require that it be directed to the most needy.

The original experimental group consisted of 108 families that moved into residence in Sumner Field Homes during the period December 15, 1938, to April 15, 1939. They were first interviewed between February 1 and April 15, 1939, and re-interviewed in 1940 from February 1 to May 31. This spread of dates over which the interviews took place illustrates another problem of field work in the community situation A zero or dead-line date at which all subjects are interviewed usually cannot be set in studies of this kind Furthermore, the control group too, consisting of 131 families who were rejected applicants to the project, were interviewed over the periods February 5 to July 31 in 1939 and February 15 to April 30 in 1940. As we have noted above, interviewing a control group involves more difficulties than interviewing an experimental group, because individuals in the latter tend to be more receptive to inquiries, whereas more resistance is encountered in the former.

How can the student bring order out of this apparent confusion of factors that present practical problems in studies made in the complicated community situation? The answer is to match the experimental group against the control group on as many known factors as pos-

[10] In the WPA study cited above, in which no time interval between measurements made before and after intruded, losses of cases comparable in size and for similar reasons were experienced See chap III, pp 41–50

sible. But matching causes additional loss of cases which cannot be matched. Thus in the housing study 82 cases were eliminated for this reason 51 from the experimental group of residents and 31 from the control group of slum families As the number of matching factors is increased the elimination of cases grows. Thus with 5 matching factors, race or culture class of husband, employment situation of husband, occupational class of husband, number of persons in the family, and income of family, some 47 cases were lost from the experimental group and 10 from the control. When 5 added controls or matching factors were introduced, 4 more dropped out of the experimental group and 21 from the control Thus the total losses of cases from the original 239 families were for all causes 157 cases, leaving 82 terminal cases that were matched on 10 factors, of which total, 44 cases were in the experimental group and 38 cases in the control group The total loss was 65 6 per cent for the one-year period of the experiment

Matching on individuals so that for each individual in the experimental group there is another in the control group alike in respect to score on every one of the matching factors results in very great losses because of inability to satisfy such a precise matching requirement In the Christiansen experiment this rigorous manual matching procedure diminished the experimental group from 671 to 23 individuals, and the control group from 523 to 23 individuals In the housing study this method of identical individual matching was not used; matching was by sub-categories. This latter method proceeds by elimination only of cases from one group which are entirely outside the categories of the other group, and is essentially that of insuring comparable frequencies in sub-categories. This method has the great advantage of avoiding numerous losses due to matching Louis Guttman in his study, "On Uses of the Critical Ratio—A Research Paper," University of Minnesota, June, 1939, pp 34, 36, has worked out the formulas used.

The next question that arises is, What sort of individuals are lost during the interval that the experiment runs? Are those who drop out because they refuse information on second interview or have

moved away different in any respects as a group from those who remain in the samples that carry through the experiment? Are those dropped for reasons of mobility and refusal different types of cases from those lost because of failure to match on factors of control?

Since we had initial measurements in 1939 on morale, general adjustment, social status, and social participation for all cases, we may examine the scores of the dropped cases for comparison with the scores of those that survived to the terminal dates of 1940. It is possible that a systematic recapture program such as used in the study of consumer purchases[11] would have diminished these losses and interpolated the measurements on some missing cases, but it would have considerably changed the dates and places of interview so that measurements on the attitude factors might have been invalidated for purposes of comparison That is, while changes in date of interview by a few weeks probably did not affect the validity of information on schedules of the study of consumer purchases and on social participation and social status in the housing study, it might have seriously changed an attitude response measured by the *Survey of Opinion* schedule in the housing study. Consequently the time factor in extending the span of time over which field work interviews are made requires careful consideration before a recapture program is organized.

In general, we found that the cases dropped, either by losses or from inability to match on control factors, were cases showing more extreme measurements, that is, higher or lower scores on all scales used to measure the effect (morale, general adjustment, social participation, and social status) Thus the net effect of losses was to increase the homogeneity of the residual groups (experimental and control) from which the results of the experiment were inferred. As a consequence of these facts the magnitude of absolute scale differences between the experimental and the control groups upon measures of effect was small and hence the critical ratios were

[11] Erika H Schoenberg and Mildred Parten, "Methods and Problems of Sampling Presented by the Urban Study of Consumer Purchases," *Journal of the American Statistical Association*, June, 1937, pp 311–322

diminished from this fact alone. Also the reduced size of the residual groups yielded larger standard errors, which, in turn, played a part in still further reducing the size of the critical ratios.

An opportunity presented itself for a check-up in November, 1942, on the original experimental group of residents. The FPHA visitors re-interviewed the families and found only 21 of the original 44 families of the experimental group still residents of the housing project. These third interviews secured 1942 measures on morale, general adjustment, social participation, and social status On attitude measurements (morale and general adjustment) the 21 families showed decided improvement over their responses in 1939 and in 1940. Whether these gains may be attributed to the effects of living in the housing project or form part of the general improvement of the times we are unable to state. We did find, however, that the remainder of the original experimental group, that had moved away from Sumner Field Homes between 1940 and 1942, showed a somewhat better morale and general adjustment in 1939 and 1940 than did the group of 21 families we were able to follow up in 1942, and who then showed such decided gains.

It will be noted that this 1942 follow-up of resident families revealed a loss of nearly 50 per cent from the original 44 families of 1939–1940 This is the most significant aspect of the follow-up considered in terms of sampling. It shows how difficult it is to preserve the conditions of an experiment for even as few as three years (e g., 1939 to 1942). If effects of social treatment are to be measured as part of rational programs for evaluating social action, how long does the experiment need to run? If a long time is desired or needed to test the effects of social treatment, then the evidence of this follow-up shows how difficult it may be to complete adequately planned experiments which are to be performed in a free community situation

We may now tentatively summarize the problems of field interview as found in three community studies using the control group technique in a free community situation

1. The longer the time of each interview (every additional scale or questionnaire to be filled in extends the time of each interview),

the fewer the cases that can be canvassed with a given size of field staff in a given time.

2 The longer the period (in days or weeks) over which the interviews run, the greater the likelihood that administrative changes will occur and disrupt the conditions of the experiment.

3 The longer the time interval over which the experiment runs, the greater the loss of cases due to mobility and refusals

4. Such losses are likely to be more numerous in the control group than in the experimental group

5 The larger the number of factors used in matching, the greater the number of cases that will be dropped.

6 The cases lost by mobility, refusals, matching, etc, tend on the average to be cases with extreme measurements on criteria of effect. Consequently the trend is toward increasing homogeneity in the experimental and the control groups with resulting diminution in magnitude of differences found on measures of effect. This is to say that losses of cases between "before" and "after" measurements tend to leave in the residual groups individuals more alike than was the case at the outset [12]

In addition to the last point it needs to be recognized that the forms of social treatment which experimental designs attempt to evaluate are usually directed upon populations originally quite homogeneous. To be specific, the kind of people who receive WPA, relief, or public housing programs are people at a relatively low and uniform level of income, occupational class, and educational attainment when the experiment begins. To this fact of original homogeneity are added the effects of losses of cases, which, as shown above, tend to create still further homogeneity Consequently if any changes have occurred it becomes evident that these can be

[12] The Hill *experiment (see chap* IV, *pp* 79–84) *confirmed some of these problems*. But the problem of selection is often present In the housing study described in the same chapter, the conditions of eligibility for admission to the project obtained among the applicants Yet these very conditions of low income, family size, ages of children, etc, made both the accepted and the deferred rejected applicants much alike Thus in so far as inference in this study is confined to the closed system of this *single experiment, the invalidating influence of selective factors is disposed of* Confirmation of results in similar experiments is, however, still much to be desired.

described only when the scales that measure effects are sensitive to small differences. Two ways in which such scales may be provided are: (a) to improve existing sociometric scales within their middle ranges by finer calibrations, or (b) to utilize the evidence of several minor changes in the same direction by means of computing multiple critical ratios This latter procedure enables the research student to measure a pattern of change in several factors.

If cost precludes the use of many samples or the use of larger samples to offset the handicaps inherent in samples attenuated from community influences, it would seem that the two procedures just noted may still supply a partial solution of both the methodological and the practical obstacles to using experimental designs which we have enumerated in this book

Finally, interference with randomization due to practical considerations of an administrative nature does not by any means invalidate the use of experimental designs in the community situation because the results do hold for the groups studied and within the limits of the known controls. The next step is to repeat the experiment on similar groups under like conditions and when possible with additional or more significant controls. The cumulative findings of several similar experiments may prove to be as useful as would fewer experiments based on control of unknowns by randomization within the sub-groups of a stratified sample.

Theoretical Limitations

Throughout this discussion of experimental designs in sociological research we have frequently used the term statistical significance as a criterion of when a difference between the means of an experimental group and a control group is important, or when a change between the means of observations at two dates is important Let us now review the meanings of these terms "important" and "significant."

When Is a Difference Significant?

In the folk language a difference is important or not important. It is important if an expert says it is But the public's judgment

of who is an "expert" is often not reliable since there is little common knowledge of technical specialization in science. Consequently importance becomes a matter of opinion. The opinion of interested parties is often taken as a criterion of when a difference between wage rates, a difference in educational achievement, or a difference in need of social treatment is important In social programs the interested party whose opinion is taken as a criterion of the importance of observed differences is sometimes the promoter.[13] But his opinion of the results of a program is naturally influenced by his *desire* to achieve some end-in-view Again, it may be the administrator of the program whose opinion of its results is taken as a criterion. But it should be remembered that the administrator has accountability and responsibility for seeing that the end-in-view *is* achieved Both promoter and administrator are tempted when making a judgment of the importance of a gain associated with a social program to substitute the psychology of *vindication* for the psychology of *verification,* to substitute the notion of justification of a position taken in advance and expressed by emotional appeal *for the operation of confirming the results* of applying a program *by repetition* of the program under like conditions. Thus is the spirit of vindication unconsciously substituted for the spirit of verification Moreover, these opinions of what is important, or when a difference or a change associated with a program is important, are rendered *after* the experiment is terminated The results are then value-judgments with an underlying means-ends schema that may be only implicit, and the whole process is subjective

To be sharply contrasted with the foregoing criterion of significance is the criterion of statistical significance, which is based on differences or changes in the magnitude of measurements. In experimental designs reliance is usually placed on the means of measurements for the sake of comparison with the means of random samples Now it may be noted that the means are derived from measurements obtained from applying some psychometric or sociometric scale. Furthermore the scale is a reliable instrument of

[13] See pp 168–169, also chap IV, p 78, footnote 22

observation which has been pre-tested for reliability before use in the study under consideration. It is *reliable* because repeated measurements by the same observer on the same or similar subjects are essentially in agreement; they show small errors of observation. It is perhaps also reliable because measurements independently made by different observers on the same subject agree. The scale is also *valid* It does measure that which it is designed to measure. There is, first, an initial validation by the selection of items to be included when specialists are asked for judgments upon these items; and there is, second, a terminal validation when the scale is shown to be consistent in discriminating differences equally well with some other scale or external and independent criterion acceptable to research specialists. All these tests are met *before* the scale is used in experimental designs. Moreover, the scale is supplemented by empirical norms,[14] or mean scores derived from its use and differentiating the two groups, so that comparisons can be made, and any new group can be placed with reference to a series of other groups whose means have been previously found.

Finally, having met all the tests just enumerated, we resort to the probability frame of reference to determine statistical significance when we find that a difference between means is three times its standard error. This is done on the basis of the following procedure. First, we know that the means of many random samples tend to approach the binomial expansion or the normal probability curve, and also that the differences between the means of random samples have the same limit. Second, we can estimate the frequency with which any magnitude of a mean or any difference between means occurs in the fluctuations of random sampling from a homogeneous and infinite universe Third, when the obtained difference is one whose magnitude can occur very infrequently in random sampling, or only once in 370 trials, we say that the difference is statistically significant, that is, it is a difference which occurs so rarely in the fluctuations of sampling from a homogeneous universe that its occurrence in the present study is *probably* indicative of a real difference between the two groups compared. Note that we cannot

[14] See Appendices A and B

say it is certainly indicative of a real difference, only that it is probably indicative of a real difference Finally, although in most cases we deal with an experimental group compared with a control group, both of which are non-random samples and hence do not conform to the strict requirements of random sampling basic to the probability theory, we can at least say that the obtained difference is greater than usually occurs in random sampling, or that it is no greater than might occur in random sampling, as the case may be. The only way in which the probability of the difference being a real one can be generalized to a universe from which the experimental and control groups were chosen is to repeat the experiment and find corroboration on successive trials Nevertheless the discovery of a significant difference by the operations outlined above is an objective procedure and supplies a degree of confidence that encourages further experimentation under similar conditions. This is its chief positive and constructive value for sociological research.

It should be remembered that the theory of sampling is used to give us a rationale for generalizations from a relatively small group (a sample) to the larger population or universe from which it was chosen The conditions are (1) randomization in the selection of the cases in the sample is essential, also the conditions (2) of a large universe (infinite), which is (3) homogeneous in composition. Faced by these conditions both the novice and the purist are inclined to insist that, unless these conditions are met or approximated with reasonable closeness in studies by experimental design, the researcher cannot formulate valid generalizations from the results of his experiment to the population at large[15] When non-random samples are chosen from a heterogeneous universe, and a limited one at that, as in the case of most of the studies described in this book, doubt is cast upon the results and a lingering sense of intellectual insecurity persists. It is true, of course, that no generalizations to any universe can be made from the results found among the non-

[15] For a more adequate discussion of this problem, see T C McCormick, "Note on the Validity of Mathematical Probability in Sociological Research," *American Sociological Review,* October, 1945, pp 626–631

random samples in the studies herein described. But then no such claim was anywhere made, we explicitly pointed out in every example that no generalization could be made from any one of these experimental studies *except* with respect to the definitive groups actually studied. We repeatedly stated that the results were limited to the closed system of each particular experiment, and that only replication of the experiment which yielded corroboratory results would supply any reliable basis for generalization to a universe. Since this was the case, of what utility was the use of probability tests on non-random samples?

Rejoinder to this question is simply that the alternative to the use of probability tests would have been to rely on subjectively formed opinions about the importance of the obtained differences or changes in each single experiment even when limited to the definitive groups studied. As was stated earlier, we reject this alternative as unsatisfactory. But suppose the difference between the experimental group and the control group, both being non-random samples, *is* statistically significant; we are then faced with the dilemma as to whether the obtained difference occurred because the two groups were chosen from different universes, or merely because the two samples were non-random. This raises a real difficulty of interpretation. In most cases the experimental group *was made* non-random, either because randomization was not possible, for the practical reasons discussed in pages 168–169, or because the social program or causal factor was brought to bear upon accessible subjects only. In any event, in the present stage of experimental work done in the free community situation we relinquish any argument for generalization that different universes are indicated even when probability tests of the differences between the non-random samples are statistically significant.

In summary, there are four considerations which seem relevant to the use of probability tests on such non-random samples as the experimental and control groups of these experiments.

1 Justification for the use of P-tests in the present provisional state of development of experimental designs in sociological research is essentially pragmatic, as when the over-all results from

the experiment are confirmed by replication,[16] despite the fact that the experimental and control groups are non-randomly selected. The P-tests serve in this context merely as a convenient device to order the data to a rational frame of reference.

2. The result of applying P-tests to non-random samples of the experimental and control groups ". . . at least allows us to say that the observed differences are no greater than might occur with random samples."[17]

3. With respect to the kind of samples, McCormick makes another point. ". . . Because of the difficulties in taking a simple sample under many conditions of sociological research, it is fortunate that the standard errors of simple samples are usually not very different from those of random samples, and in any case are somewhat larger For these reasons, investigators often apply simple sampling errors to random or even stratified samples, in order to save labor or to be on the conservative side when in doubt as to what the error formula should be."[18]

4. The principle which lies at the basis of P-tests used in the present stage of experimental design in sociological research is. the frequency *distributions of the means* of measures of items in large random samples tend to approach a normal probability distribution, *whatever* may be the distribution of frequencies of measures in the universe, and similarly for the *differences* between the means of random samples. Empirical evidence supporting this principle is widely accepted by statisticians. Moreover, the evidence also tends to approximate even more directly the theoretical model of the binomial expansion. This is a principle which any student may confirm empirically by the tossing of coins, or by placing in a frequency distribution the means of some measured trait of many random samples.

We may now summarize this discussion of the application of

[16] It is unfortunately still the fact that replication of experimental design studies in sociological research has been infrequent Therefore this statement is one of expectation rather than of fact For further discussion of this point, see footnote 22, p 78, and pp 187–189 See especially chap VIII

[17] T C McCormick, *Elementary Social Statistics*, McGraw-Hill, 1941, p 256

[18] *Ibid*, p 234

P-tests to random and to non-random samples in the schematic fashion shown in Table 27. To make an application of this scheme, let us take the data of the housing study of Chapter IV, pages 72–73, using the multiple critical ratios for the differences, changes, and

TABLE 27 Schematic Form of Experimental Design for P-Tests

[a] The heuristic equations in col (5), expressing changes (D) in the magnitude of means on some measure of effect, should yield a critical ratio of 2 at least in row 4, col (5), and preferably 3 Similarly the differences in row 5, col (4), and the differences between changes in row 6, likewise

It will be noted that these requirements are met in several of the experiments described in this book See chaps IV, pp 70–74, and V, pp 122–124

$$\text{Critical ratio} = \frac{(\bar{X}'' - \bar{X}')}{S\,D_{(\bar{x}'' - \bar{x}')}}$$

$$\text{Multiple critical ratio } \frac{Cr_1 + Cr_2 + \cdot + Cr_m}{\sqrt{m}}$$

differences in change. The multiple critical ratio obtained for column (3), row 5, is found to be 0.18, which is as it should be if the condition is met that the two groups began the experiment in 1939 essentially similar in respect to the pattern of three measures of the effects of housing For 1940, the multiple critical ratio *increases* to 1 65, so that although the second condition is not precisely met,

it is nevertheless approximated, since the difference in 1940 approaches statistical significance. When, however, the other criteria are checked, the conditions are met. Thus the multiple critical ratio of column (5), row 3, is 1 82, and of column (5), row 4, is 6.01, and the chief criterion of the significance of the comparison, the difference between changes in the experimental group and the control group, yields a multiple critical ratio of 4 97 The reader may also wish to turn to Table 26, page 133, of the rental-tuberculosis study, as another example of the utility of this tabular form as a device to check results.

Although we make no attempt in this book to deal with the detailed mathematical-statistical aspects[19] of the problem created by the application of probability tests to such non-random samples as an experimental and a control group, but attempt merely to state in verbal terms the logic of the problem, it may be helpful to note briefly a semantic confusion which sometimes grows out of an uncritical use of the probability theory, which, if not stated explicitly, tends to embarrass our reasoning rather than to clarify it. Confusion results when an indeterminate and a metaphysical meaning of probability is accepted, and clarification results when an empirical meaning is taken.

Logicians have noted the fact that there are at least two meanings[20] and perhaps three meanings[21] of probability in the literature. Most students of mathematical statistics,[22] and following them the authors of statistical textbooks,[23] *use the empirical meaning of probability as the measure or limit of relative frequency.* It is this latter meaning that has significance for the student of experimental designs and is the meaning used in this book.

[19] M G Kendall, *The Advanced Theory of Statistics*, Lippincott, vol 1, 1943

[20] R Carnap, "The Two Concepts of Probability," *Philosophical and Phenomenological Research,* June, 1945, pp 513–532

[21] E Nagel, *Principles of the Theory of Probability,* University of Chicago Press, 1939

[22] Kendall, *op cit*, and R von Mises, *Probability, Statistics, and Truth,* 1928 and 1939.

[23] Margaret J Hagood, *Statistics for Sociologists,* Reynal & Hitchcock, 1941, see especially pp 354–369; and T C McCormick, *Elementary Social Statistics,* McGraw-Hill, 1941, pp 234, 256

The classical conception of probability of Bernoulli and Laplace[24] and of De Morgan,[25] according to which all of our knowledge is regarded as having a probable character, and hence a degree of probability is merely a measure of the amount of certainty associated with belief, is not accepted by Carnap, Nagel, and von Mises. Nor do they accept the meaning of probability as that of a unique logical relation, as used by Keynes[26] and Jeffreys.[27] As to the former, Carnap states, "It must be admitted that the classical Laplacean definition is untenable";[28] and Nagel holds that it is not clear whether, in this meaning of the concept, probability is regarded as *a measure of belief* (psychological) or as a measure of the degree of belief one *ought* to entertain as reasonable, and finds evidence of circular logic in the Laplacean definition[29] As to the second conception of probability (Keynes, Jeffreys), Carnap says, "Modern authors, especially Keynes, Jeffreys and Hosiasson[30] have proceeded more cautiously, *but at the price of restricting themselves to axiom systems* which are rather weak and hence far from constituting an explicit definition"[31] (italics ours); and Nagel states, ". . . the conception of probability according to which we cannot in principle control by experiment and observation the probability statements we make is not a conception which recommends itself as germaine to scientific inquiry"[32]

The first two conceptions of probability have been reduced by Carnap to one: the degree of logical confirmation, which he designates as P_1. He states that it is "very important to distinguish clearly between *kinds of events* (war, birth, death, throw of a die, throw of this die, throw of this die yielding an ace, etc.) and *events* (Caesar's death, the throw of a die made yesterday at 10 A M., the series of all

[24] P Laplace, *Essai Philosophique sur les Probabilités,* 1814

[25] A de Morgan, *An Essay on Probability,* 1838

[26] J M Keynes, *Treatise on Probability,* 1921.

[27] H Jeffreys, *Scientific Inference,* 1931

[28] Carnap, *op cit*, p 519

[29] Nagel, *op cit*, p 46

[30] J Hosiasson-Lindenbaum, "On Confirmation," *Journal of Symbolic Logic,* 1940, pp 133–148

[31] Carnap, *op cit*, p 520

[32] Nagel, *op cit*, p 50

throws of this die past and future),"[33] and continues, "It is clear from probability statement$_1$, a statement on frequency can never be inferred, because the former is purely logical while the latter is factual. . . ."[34] The second meaning of the concept of probability is empirical—the limit of relative frequency. Of this meaning Kendall says, ". . . The second approach seeks to define probability in terms of the relative frequency of events and thus to throw the theory back to pure mathematics of abstract ensembles (Kolmogoroff, 1933) or to the limiting properties of sequences (von Mises, 1936) . . .";[35] and again, ". . . a considerable measure of agreement that the concept of probability is founded on our experience of the frequency of observed phenomena . . ."[36] exists among students of the subject. All of which brings us back to the contention that when probability tests are used as illustrated in the studies of this book, *to serve only as an empirical safety device,* and not to provide a basis for extensive generalization, then their use is justified and should help dispel the doubts of the novice and the criticisms of the purist

A schematic presentation of our analysis appears in Table 28 and shows the kind of referents which seem to stand for the three different meanings of probability. Of the three meanings only the third, row 3, columns (2) and (3), has any fact referent. For this third meaning—probability is the limit of relative frequency—we have at least two kinds of empirical evidence. The one (1), may be obtained by the reader from tossing coins, let us say 6 pennies tossed 64 times, recording after each toss the number of heads up, and then comparing the frequencies with the coefficients of the binomial expansion, $(\frac{1}{2}+\frac{1}{2})^6$. Several such experiments will yield frequencies that will fluctuate in magnitude about the magnitudes of the coefficients of the binomial expansion and, in the long run of many repetitions of the experiment, will approach the binomial expansion as a limit The other (2), may be obtained by the reader by selecting random samples of uniform size from some reasonably large universe, and then computing the means of some measure

[33] Carnap, *op cit*, pp 522–523
[34] *Ibid*, p. 526.
[35] Kendall, *op cit.*, p. 165.
[36] *Ibid.*, pp. 165–166

and arranging these means in a frequency distribution of seven classes. If several groups of random samples are analyzed in this way and the resulting frequency distributions compared with the coef-

TABLE 28 Probability and Its Referents [a]

Concept of Term	Referents	
	Having Linguistic Referents (*Designata*)	Having Fact Referents (*Denotata*)
(1)	(2)	(3)
"Probability"	1 A measure of belief (1) degree of belief (intensity)? (2) degree of belief one *ought* to entertain?	(?)
	2 A unique logical relation which cannot in principle be controlled by observation or experiment	(?)
	3 The measure or limit of relative frequency (an operational definition) →	(1) Tossing dice or coins many times yields a frequency distribution of attributes which corresponds to the binomial expansion (2) The means of random samples tend to approach the binomial expansion or the probability distribution

[a] Although this table is similar to Table 1, pp 23–24, it will be noticed that the order of the second and third main columns is reversed (for logical reasons) in this case

ficients of the binomial expansion noted above, it will be found that the distribution of the means tends to fluctuate about the binomial and approach it as a limit. In this manner the student may perform certain operations and obtain confirmation of the definition of probability as the limit of relative frequency.[37]

[37] For a more adequate discussion, see Hagood, *op cit*, pp 354–358, 366

The Null Hypothesis Again

Our earlier discussion of the advantages in sociological research of the null hypothesis (Chapter IV, pp 70–74) need not be repeated here, but it may be worth while nevertheless to note that the use of probability tests[38] is facilitated by the null hypothesis, since the relevant evidence to support or to refute this hypothesis is based, in part at least, upon the relative frequency of occurrence of various magnitudes of difference or of change. In the closed system of a single experiment the null hypothesis is particularly helpful. The positive type of hypothesis in sociological research usually posits "improvement," or "change" in a desired direction, and is thus an open-ended statement, one which has an indefinite end approached only by an infinite number of small increments, and an end that often recedes as it is approached (the higher the standard of living goes, the more ways there are of being poor!). An open-ended system of effects is difficult to handle empirically. But the null hypothesis, by comparison, provides a closed system Departures from an arbitrarily fixed zero point can be measured.

Obstacles to Replication

Confirmation of the results of tests of some program by experimental design is the criterion[39] we have set up as the goal of ex-

[38] For further discussion of the null hypothesis see Hagood, *op cit*, and Lindquist, *op cit*, pp 15–16, where he remarks, "In many sampling studies the interest is not so much in the limits within which a parameter may confidently be said to be, as in the single possibility that the parameter is zero such hypotheses —that the parameter is zero—are known as null hypotheses If a statistic is such that the null hypothesis may be rejected with confidence, we say that the statistic is significant, meaning that it signifies that the parameter value is not zero however . . to prove the difference significant does not establish the cause of the difference . ."

[39] Says Alfred J Ayer, "The criterion which we use to test the genuineness of apparent statements of fact is the criterion of verifiability We say that a sentence is factually significant to any given person, if, and only if, he knows how to verify the proposition which it purports to express—that is, if he knows what observations would lead him, under certain conditions, to accept the proposition as being true, or reject it as being false. . . ." *Language, Truth and Logic*, Oxford, 1936, pp 19–20.

perimental study. Why is repetition of the experimental test of a hypothesis so infrequent in sociological research?

There seem to be several explanations of this state of affairs. Experimental designs are new and strange They run counter to certain established predilections which are current in social research. In fact, it was as recent as 1932 that many sociologists asserted that there never could be any experimental study of human relations.[40] Philosophical controversies about the meaning of the concept "cause and effect," together with the ethical complexities of the means-ends schema, effectively combined to confuse discussion of experimental designs in sociological research. Existent results of crude experimental study were often indecisive and hence seemed additional substantiation of the claims of its critics. Some students may have been reluctant to undertake systematic and planned research that required thorough analysis and the use of measurement devices The expense of an experimental study may also have been a deterrent. In connection with this last point it may be of interest that the fairly extensive studies of morale and relief of Chapter III, pages 41-50, and of the social effects of public housing, Chapter IV, pages 58-79, cost in each case the sum of approximately $2000 to complete, from planning through publication, and including supervision.

Now that many standardized sociometric scales are available for social observation and measurement, and a variety of exploratory efforts at experimental study have defined the patterns of the projected and the ex post facto experimental designs, it would seem that the situation in sociological research is ready for repetition of some of the pioneering studies. The fact that these investigations were far

[40] Robert C Angell, "The Difficulties of Experimental Sociology," *Social Forces*, vol XI, No 2, December, 1932, pp 207-210, lists the difficulties as recognized in 1932 In general, confusion of thought about experimental sociology stems from failure to distinguish between (a) experiments in social reform by the method of trial and error, which inevitably include *restraints* on certain individual freedoms, and (b) experimental design study that makes observations of human relations under conditions which control some of the variable factors, not by interference with individual freedoms, but by matching the measurements of several factors See also, H C Brearley, "Experimental Sociology in the United States," *Social Forces*, December, 1931.

from perfect experimental designs is no excuse for the failure to try to repeat them A later study of public housing in Newark[41] made during 1942 and 1943 did not utilize the standardized sociometric scales which had been used in the Minneapolis study described in Chapter IV, although the results of this Minneapolis study had been published in December of 1940[42] If the techniques of experimental study of the problems of human relations are to be made more reliable and precise, they should be tried out again and again. Repetition in the use of such tools of social observation as many of the sociometric scales described in Chapter VI has led to great improvement in some of these devices. It remains for the student of human relations to repeat also in the use of experimental designs for sociological study.

The Value of a Verification of Values

John Dewey contends that "operational thinking needs to be applied to the judgment of values just as it has now finally been applied in conceptions of physical objects Experimental empiricism in the field of ideas of good and bad is demanded to meet the conditions of the present situation."[43] And again, ". . . carrying over experimental method from physics to man concerns the import of standards, principles, rules" "They (the rules) would lose all pretense of formality. . . The change would do away with the intolerance and fanaticism that attend the notion that beliefs and judgments are capable of inherent truth and authority." ". . Any

[41] J Rumney and S Shuman, *The Social Effects of Public Housing in Newark, N J*, Housing Authority of the City of Newark, November, 1944, 95 pp, 2nd printing, March, 1946 The authors state on p 2, "It was possible for us to match the two groups only on the basis of certain general characteristics," but do not make clear to the reader what these factors are There is some internal evidence in the report that the factors age, sex, and race were used (pp 24–25, 28–29, 32), and yet when the published tables are examined it is evident that race (Negro, white) was far from equalized in comparisons Furthermore, the authors state (pp 38, 43, 46, 69) that the obtained differences were not statistically significant but do not inform the reader as to exactly what the differences or critical ratios were, nor whether the criterion was the 2-per-cent or the 1-per-cent level See also, for comment on this study, chap iv, footnote 22

[42] *American Sociological Review*, December, 1940, pp 868–879

[43] *The Quest for Certainty*, Minton, Balch & Co, 1929

belief as such is tentative, hypothetical." "A moral law . . . its soundness and pertinence are tested by what happens when it is acted upon," and again, "The test of consequences is more exacting than that afforded by fixed general rules."[44]

These are encouraging words and stimulating ideas. It hardly seems necessary to point out that the experimental studies described in this book (except the tuberculosis study of Chapter V) are tests of the efficacy of specific means to attain desired ends, in short, they are examples of operational thinking about social values. This is particularly the case for the six studies which concern adjustment as an end to be achieved by such means as programs of public housing (pp. 41-50), social treatment of juvenile delinquency (pp. 84-91, 97-99), and formal classroom education for the masses (pp. 99-124) or avocational counseling of individual students (pp. 79-93). Is it too much to hope that experimental designs in sociological research as exemplified in these pioneering studies may have blazed the trail that will enable the student of human relations to accomplish two much-to-be-desired goals first, to appraise by scientific methods of research the effectiveness of specific means to achieve certain ends; and second, to be able to isolate cause-and-effect relationships in the complicated area of human relations?

[44] *Ibid*, pp 277-278.

Chapter VIII[1]

ANALYSIS OF VARIANCE AND THE *t*-STATISTIC: UNDERLYING ASSUMPTIONS

THE classical statement of the logic of experimental method of science runs something like this "Experiment, the great weapon of scientific inquiry, derives from the ability of the experimenter to replace . . complex systems of causation by simple systems in which only one causal circumstance is allowed to vary at a time. This is perhaps an ideal, but it is one which is closely approached with the technique of modern laboratory practice"[2] On the first page of the present book such a point of view and the reasons for it are outlined.

In many fields of inquiry, especially in the psycho-social sciences, laboratory techniques are difficult to develop and to apply[3] Furthermore, ". . . If the organism is taken out of its setting into the laboratory, the problem is ruptured"[4] Finally, as knowledge ex-

[1] For a key to periodical abbreviations used in chapters VIII, IX, X, and XI, see list in Appendix E

[2] G U Yule and M G Kendall, *An Introduction to the Theory of Statistics*, Griffin, London, 1940, p 2 W J Dixon and F J Massey, Jr, *Introduction to Statistical Analysis*, McGraw-Hill, 1951, p 119

[3] The ground plan of the laboratories for social research at the University of Minnesota was developed by F Stuart Chapin, and the technical equipment for experimental study was developed by Ned A Flanders and John G Darley For recent developments in laboratory planning and equipment for experimental work in the social sciences, see Robert F Bales and Ned A Flanders, "Planning an Observation Room and Group Laboratory," *A S R*, December, 1954, pp 771–781 An accessible example of a successful use of the Minnesota laboratories for small group research will be found in Reginald A H Robson, *An Experimental Study of the Effects of Different Types of Communication on Morale in Small Groups*, a Ph D dissertation, University of Minnesota, 1951.

[4] Palmer O Johnson, "Modern Statistical Science and its Function in Educational and Psychological Research," *Scientific Monthly*, pp 385–396 (especially p 387), June, 1951, W B Cochran and G Cox, *Experimental Designs*, Wiley, 1950, D T. Campbell, "Designs for Social Science Experiments," Behavioral Science Study Group, University of Chicago, 1953, W B Cochran, "Analysis of Records with a View to Their Evaluation," Milbank Memorial Fund, New York, 1954, pp 228–237

pands, the number of variables to be controlled increases with complexities of interrelationships revealed by research. How, then, is the social scientist to study problems of human relations in situ and as part of a web of fluctuating influences in the community?

Mathematical statisticians point out that R A Fisher's analysis of variance and co-variance,[5] invented by him in 1925 and now widely used in the biological sciences, offers efficient solutions of these problems by: (1) investigating an experimental problem when several variables are undergoing change concurrently; (2) doing this in the life situation without formal laboratory controls over disturbing variables; and (3) accomplishing in one continuous series of operations, and with greater efficiency (because without matching no cases are lost, and *all* the information is utilized), that which several critical ratios or *t*-tests would do Since this method of experimental design has been so successful in studies of lower organisms, it is suggested that it should be applied in research on human relations so that these studies also may be relieved of the incubus of strict control by laboratory technique By contrast with the power and efficiency of analysis of variance, it is evident that the control group studies described in the preceding pages leave much to be desired.

Let us, then, examine some empirical aspects of an application of analysis of variance to the kind of problem described in the preceding pages. Analysis of variance makes two kinds of assumptions: (1) that the observations of the effects of treatment are objective or more often exact and precise. For example, in the field of agricultural biology we have such measures as acres, age in months, bushels, calories, grams, etc.; and (2) certain underlying mathematical assumptions have to be met.

The assumption of objective, exact, and precise measures is discussed in some detail in Chapter VI, "Sociometric Scales Available for Control and the Measurement of Effects," pages 140–164. In Chapter XI, "Some Problems of Psycho-Social Measurement," some new scaling methods are described as well as the problem of guarding against falsification of responses. Since psychometric and socio-

[5] R A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd, 1925, and R A. Fisher, *The Design of Experiments*, Oliver & Boyd, 5th Edition, 1949

metric instruments are sometimes fallible, the conditions of objectivity, accuracy, and precision are often not strictly satisfied

The mathematical statistical assumptions which underlie valid use of analysis of variance are several: (1) the group to be studied should be a random sample[6] from a large universe (ideally an infinite universe) which is defined in advance[7] This provides assurance of independence[8] among the measurements on the cases, a pre-condition to experimental study of the effects of treatment In other words, independence means that every member of the universe or population sampled should have an equal chance of being included in the sample; (2) the cases (with the measures on them) are drawn from normally distributed populations, (3) all groups or samples have the same variances (homoscedasticity); and (4) the means of these normal populations are linear combinations of "effects," due to column and/or row influences, i e., effects are additive.

In general, it is difficult, if not impossible, to satisfy the foregoing mathematical statistical assumptions in designs for research on human subjects in the free community situation This is the chief reason why analysis of variance was not used in the studies described in this book In lieu of analysis of variance, use was made of the critical ratio (pages 42, 46, 120) or the *t*-statistic for estimating the significance of differences between groups; and in lieu of control by randomization, resort was made to matching[9] on several variables known by previous research (page 61) to be related or for which there were reasonable grounds calling for control But let us return to a point-by-point consideration of the practical difficulties to be overcome if the mathematical statistical assumptions which underlie

[6] Dixon & Massey, *op cit*, pp 126, 133, 138, 141, and 147 See also W G Cochran, F Mosteller, and J W Tukey, "Principles of Sampling," *J A S A*, March, 1954, pp 13–35

[7] Definition of the universes of psycho-social measurements presents problems to be discussed in Chapter XI

[8] Palmer O Johnson, *Statistical Methods in Research*, Prentice-Hall, 1949, p 187

[9] Quite apart from the usefulness of matching to attain control over variable factors, which is, under appropriate conditions, more adequately achieved by randomization, matching has empirical science values in the sense of contributing to a more thorough knowledge of one's data This point was discussed on pages 110–111, in footnote 17

valid use of analysis of variance in the context of the free community situation are to be met Subsequently the limitations on use of the *t*-statistic will be considered.

In the first place, random selection of human beings to receive or not to receive some needed treatment or service is seldom possible in the community setting This problem of obstacles to randomization is discussed at length on pages 165–169 and need not be repeated here except to describe one solution which I proposed in 1950.[10]

When the sociologist attempts to apply an experimental design to measure the effects of some social program of treatment, such as a public housing program in the free community situation, two obstacles to generalization of results are encountered: (1) the selection of slum families for an experimental group to receive treatment and a control group denied this treatment (admission to the public housing project) cannot ordinarily be randomized, because administrative rules always require that families to be admitted to the project must be the *most needy;* and (2) during the experimental period of one to five years, there are losses of cases due to death, mobility, etc, which would destroy any randomization at the outset In one study of this problem these losses of cases due to death, illness, mobility, refusals, etc, amounted to 12 per cent of the experimental group of slum families admitted to the housing project, and to 42.7 per cent of the families in a matched control group remaining in the slum for the one-year run of the study. The following design of experimental study could obviate the first difficulty

The operations would be: (1) take a housing project of limited accommodations, say, 500 dwelling units; (2) build up a pool of 1000–1500 eligible and processed families who could be admitted; (3) explain to applicants for admission to the housing project that the limited accommodations require that applicant families draw lots for admission (randomization); (4) then the families that drew

[10] In my original research note I proposed a solution of the second problem of "lost cases" by means of a sub-sampling procedure This proposal is, however, subject to difficulties of interpretation of results In correspondence with Harold Hotelling and Nathan Keyfitz it was pointed out by them that there were theoretical difficulties with my proposal Therefore, this second part of my original proposal is not reproduced here See F Stuart Chapin, "Research Note on Randomization in a Social Experiment," *Science,* December 22, 1950, p 700

lucky numbers will be admitted; and (5) the families that drew unlucky numbers will have to wait their turn as further construction opens up new projects. These rejected families become the control group remaining in slum conditions. In this manner favoritism and bias in admission would be avoided and yet randomization would be obtained Both groups would be measured for adjustment at the beginning of the experiment, followed through an experimental period of one to five years, and then measured for adjustment at the terminal date.

The second dilemma is loss of cases from death, illness, mobility, refusals, etc, during the run of the experiment, thus destroying the initial randomization.[11] The resolution of this dilemma is a more difficult problem but would appear to be found in further developments along two lines of procedure: (1) Alfred Politz[12] and Willard Simmons propose a method of obtaining information about "not at homes" and of getting these into the sample without callbacks, and (2) the literature on "missing plots"[13] provides procedures to deal with such problems which may be adapted to the problem of "lost cases" in psycho-social studies of human relations. Fisher's method of dealing with "missing values" is a case in point,[14] although it applies to the procedures of regular analysis of variance.

Yet even had randomization been possible, with its advantages of independence, control, and representativeness, it is never quite certain whether random selection of human subjects can yield strict independence It was pointed out by L R Hafstad[15] in 1940 that in the social sciences, observations are not ordinarily independent in

[11] Marvin J Taves supplies empirical evidence on randomization and offers an alternative solution of the problem in his article, "An Experimental Design to Preserve Randomization in Social Experiments," *A.S R*, February, 1953, pp 90–96

[12] "An Attempt to get 'Not at Homes' into the Sample Without Callbacks," *J A S A*, March, 1949, pp 9–31

[13] C Radhakrishna Rao, "General Methods of Analysis for Incomplete Block Designs," *J A S A*, December, 1947, pp 541–561, especially section (c), pp 543, 561, also R C Bose and T Shemamoto, "Classification and Analysis of Partially Balanced Incomplete Block Designs in Two Associated Classes," *J A S A*, June, 1952, pp 151–184

[14] R A Fisher, *The Design of Experiments*, pp 175–181, see also Palmer O Johnson, *Statistical Methods in Research*, pp 105–107

[15] L R Hafstad, "On the Bartels Technique for Time-Series Analysis and Its Relation to Analysis of Variance," *J A S A*, June, 1940, pp 347–361 (Part I)

the strict sense an economic index for any one day, month, or year, is dependent on that of the preceding time unit; the number of strikes in progress is dependent on the number in some previous unnamed time, etc. In the context of our present discussion of method, statistical independence among cases selected for study in biological experiments with plants or animals is not as difficult as with human beings, since even when the nominal requirements of statistical independence are met there is still the chance in social researches that some eloquent or plausible person will influence the response to treatment of the other members of a sample group by his arguments or conversation. In the case of such groups as juries this is, of course, notoriously true. Since verbal stimuli and verbal responses are often taken as measures of response to treatment, as in the scales that measure morale, general adjustment, and the like, it is evident that samples composed of human beings, however randomized, present problems of measurement, estimation, and inference that are complex and difficult to interpret in terms of the condition of strict statistical independence

As to the other underlying assumptions of mathematical statistics, there are also difficult problems in the application of analysis of variance in studies of human relations in the free community situation. The assumption of normal distribution of measures can seldom be satisfied, although several measures of effect, such as morale,[16] general adjustment,[17] social status [18] (as used in the WPA-Relief and Housing Studies, pages 41–50 and 61–65), show symmetry of distributions approaching normality, or for rentals slight skewness (pages 124–133) [19] On the other hand, social participation shows a *J*-curve and its sampling distribution is somewhat like χ^2

[16] E A Rundquist and R F Sletto, *Personality in the Depression*, University of Minnesota Press, 1936, pp 45, 113

[17] *Ibid*, pp. 223–224

[18] The means of 41 samples show a distribution approaching normality with a χ^2 test of goodness of fit at $P_{20} > P > P_{10}$ for 5 degrees of freedom

[19] C F Schmid, *The Social Saga of Two Cities*, Minneapolis Council of Social Agencies, 1937, for the Twin Cities as of 1934, p 210, and an unpublished sampling distribution of 50 random samples of census tracts, drawn from 721 census tracts of Brooklyn, N Y., for 1940, also C F Schmid, *Social Trends in Seattle*, University of Washington Press, 1944, p 234, data for 1940

distribution for five degrees of freedom.[20] Both Dixon and Massey,[21] and also Cochran,[22] show that moderate violations of the assumptions of normal distribution and equal variances change the results of analysis very little It is therefore not so much that psycho-social measurements fail to satisfy one assumption of analysis of variance as that they tend to violate two or more of the assumptions,[23] which supplies the reason for not using analysis of variance in this book Furthermore, since in all these studies only two groups were compared, the experimental group and the control group, or as in self comparisons of "before" and "after," the t-test was used. The reason is that when only one difference is to be tested, analysis of variance, with its numerous assumptions to be met, is no economy in effort although of higher efficiency. In fact, if there are only two groups to be compared, the F-test becomes the algebraic equivalent of the t-test,[24] since $F = t^2$

Empirical research workers who desire a more systematic discussion than the foregoing of the consequences of violating the underlying mathematical statistical assumptions of analysis of variance, will find it worthwhile to read the entire issue of *Biometrics* for March 1947, which contains papers by Churchill Eisenhart, William G Cochran, and M S Bartlett. In his paper, Eisenhart shows that the assumptions of analysis of variance are less arduous if the purpose of the research is only to estimate the mean differences and merely to summarize the properties in hand (which he designates as Class I, model I), rather than to draw statistical inferences to larger populations (Class II, model II—the real purpose of analysis

[20] The means on social participation for 41 samples showed a distribution similar to the χ^2 distribution for 5 degrees of freedom

[21] *Op cit*, p 126

[22] W G Cochran, "Some Consequences when the Assumptions of Analysis of Variance Are Not Satisfied," *Biometrics*, March, 1947, pp 22–28

[23] Especially the assumption of linear correlation, since many measures of psycho-social responses show curvelinear relationships See M Ezekiel, *Methods of Correlational Analysis*, Wiley, 2nd ed, 1941, pp 75–127, for an explanation of curvelinear correlation analysis by free-hand curves

[24] G W Snedecor, *Statistical Method*, Ames, Iowa, 1937, pp 185, 200, 259, and E F Lindquist, *Statistical Analysis in Educational Research*, Houghton Mifflin, 1940, p 98, also A L Edwards, *Experimental Design in Psychological Research*, Rinehart, 1950, pp 184, 277

of variance). An interesting example of Eisenhart's Class I appears to be J C Brown's "An Experiment in Role Taking" [25] Use of analysis of variance for the limited purpose of merely summarizing data seems to be the answer to Margaret Hagood's query, "Are the methods of analysis of variance (and co-variance) applicable for observational situations where data are observed in the cross-classifications in which they are found rather than in the cross-classifications in which they have been randomly placed as in an experiment?" [26]

Since the critical ratios or the *t*-statistic were used in studies of this book, it now becomes advisable to examine the underlying assumptions for valid use of the *t*-statistic on paired or matched groups in experimental designs which are used on human beings in the free community situation.

John E Walsh [27] states three assumptions made in the use of *t*-statistics in experimental design (1) the totality of observations are statistically independent; (2) they all have the same variance; (3) they are all normally distributed.

These are in addition to the mathematical model,

$$y_{ijk} = \mu + r_i + \beta_j + e_{ijk}$$

Since we have discussed above the problem of obtaining strict independence among observations by means of randomization in the selection of cases for treatment, it will be sufficient at this point merely to indicate the extent to which random groups were used in these reported studies Cases were randomly selected in the Mandel Boy Scouts study and in the Hill participation study; but cases unavoidably lost destroyed randomization. Random samples or a random sub-sample (for the sake of comparison) were used in the WPA-Relief, the Christiansen school study, and in the Shulman and T B studies. No random selection was made in the Dodd rural

[25] "An Experiment in Role Taking," *A S R*, October, 1952, pp 587–597

[26] Margaret J Hagood, *Statistics for Sociologists*, Renal & Hitchcock, 1941, p 585, and revised edition by Hagood & Price, Holt, 1952, p 401

[27] "Some Non-Parametric Tests for Student's Hypothesis in Experimental Design," *J A S A*, September, 1952, pp 401–415, see also Dixon and Massey, *op cit*, pp 99, 103, 107, and 119

hygiene study or in the Minneapolis housing study; and the Barer study of juvenile delinquency was a self-comparison.

It should be noted, however, that in several studies in which random selection was incomplete or lacking (WPA-Relief, Minneapolis housing, and T.B), several of the measures used were symmetrically distributed or approached normality. Footnotes 16 to 20 present the evidence Thus there was some degree of compliance with Walsh's three conditions. We do not claim that this compliance was mathematically exact but in some cases the approximation was sufficiently close to support our use of the *t*-statistic Furthermore, it may be noted that most of our samples were not very small, thus yielding more conservative estimates by the *t*-test.

The Cochran-Cox test[28] of an hypothesis of equality of means with no hypothesis about the population variance when $N_1 \neq N_2$, $s_1 \neq s_2$, is now available for the type of *t*-test that can be validly used in comparing two groups

In concluding this chapter, where do the foregoing analyses lead us with respect to valid use of the critical ratio or *t*-statistic, to estimate the probability of occurrence in chance among hypothetical random samples of the same size as the non-random sample groups (used in the studies reported in this book) ?

Scientific validation of these empirical experimental designs, with appropriate modifications, will, it may be hoped, develop in the future toward the objective of analytical proof Examples of developments in this direction by Guttman in his method of scale construction, and by Haberman in his method of partially ordered systems of expectations applied to control group comparisons, will be described in Chapters XI and IX respectively Meanwhile it may be noted, in passing, that despite the well-founded argument of biometricians for the necessity of analytical proof designed to place the procedures of empirical experimental design on a genuine scientific basis instead of remaining only on an empirical level (an argument

[28] P O Johnson, *Statistical Methods in Research*, pp 74–75, Dixon and Massey, *op cit*, pp 147, 340, J E Walsh, "On the Information Lost by Using a *t*-Test when Population Variance Is Unknown," *J A S A*, March, 1949, pp 122–125

to which we agree), it may be pointed out that even in careful compliance with the requirements of mathematical models there is still an element of analogical reasoning involved

The existence of an analogical element attached to an otherwise analytical process comes about in this way (and it does not, of course, invalidate the scientific mathematical proof) · whenever empirical frequency distributions are available, as for example the fact that the means or differences between means of many random samples [29] approach the ideal or normal probability curve,[30] it is standard practice to regard such empirical distributions *as if* they were normal (or to test the degree of departure from normality), and then to use this analogy in justification of taking the standard errors of these sampling distributions to enter the formal probability tables [31] to determine the expectation of occurrence of observed numerical values as chance events Experience has justified this procedure, although it is clearly a form of logic resting on an analogy between the ideal (mathematical and analytic) and the empirical Thus, there is a fortunate conjuncture of the empirical and the analytical without which the whole procedure would be pointless. When non-random samples drawn from populations known to be different, also differ markedly on a measure of practical significance for the problem in hand, then it would seem that the critical ratios or *t*-values should differ also significantly That this can, in fact, be the case is shown on pages 242–244 This is no proof, of course, that the reverse would be the case; namely, that a high *t*-test between two non-random samples would per se identify them as chosen from different universes. In the studies reported in this book, whenever the distributions were not normal or unknown, and especially when the *t*-statistic was computed from non-random samples, the analogy factor was compounded. Besides the primary analogy described above, there was also present a secondary analogy. It occurs in this way. The *t*-formula when applied to non-random samples depends

[29] Johnson, op cit, pp 31–61

[30] When they satisfy certain standard tests for normality, see Dixon and Massey, *op cit*, pp 50, 61–62, and *passim*

[31] *Ibid*, pp 295–361

on the argument: *if* we had random samples of this size and variance, *then* we might expect these estimates of the parameter value (the mean in these cases) to occur by chance this many times in theory. When so used, t-tests have no mathematical base; their only utility is to serve as a device for rough comparison by analogy to what *could* have happened as a chance event.[32]

[32] See also pages 181 to 182 above

Chapter IX

NON-PARAMETRIC OR DISTRIBUTION-FREE STATISTICAL METHODS

In the preceding chapter, the conditions of valid application of analysis of variance and the t-statistic were briefly discussed In the present chapter brief consideration is given to some alternative procedures which obviate the normality assumptions about the distribution in the universe that is sampled. Lincoln E Moses,[1] in his useful summary of non-parametric methods for psychological research, cites Poincaré's famous remark that, "Everybody believes in the law of errors (the normal distribution), the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact."[2] Moses then proceeds to note that until recently the only distribution-free methods were rank order correlation and Chi-square tests[3]

It is worthy of note that John E Walsh states, "Sufficient conditions for the validity of non-parametric tests and confidence intervals are (I) there is statistical independence among the blocks; (II) within the same block, linear combinations of treatment effects (i.e. the observations yielded by the treatments) have probability distribu-

[1] Lincoln E Moses, "Non-Parametric Statistics for Psychological Research," *Psy Bul*, March, 1952, pp 122–143, see also Keith Smith, "Distribution-Free Statistical Method and the Concept of Power-Efficiency," in Part IV, *Research Methods in the Behavioral Sciences*, edited by Leon Festinger and Daniel Katz, Dryden, 1953

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1941, p 232

[3] A E Treloar, *Elements of Statistical Reasoning*, Wiley, 1939, pp 210–226, supplies an unusually clear and logical exposition of the Chi-square test Although the notation used is "classical," and differs from that of the "modern," it is for this very reason unusually cogent Ordinarily the χ^2 test on less than 50 cases loses effectiveness, therefore the tables of E Federighi in "The Use of Chi-square in Small Samples," *A S R*, December, 1950, pp 777–779, will be found useful for totals of 40 or less cases G U Yule and M G Kendall, *An Introduction to The Theory of Statistics*, Griffin, 1940, gives on pages 422–423 a concise statement of conditions for use of the Chi-square test

tions (cumulative) which are continuous and symmetrical If care is exercised in setting up the design, it appears that situations where these assumptions hold to a reasonable approximation can usually be obtained in practice. . . ."[4] It will be seen that condition I is the form of independence required by Chi-square, and that variations from strict linearity may occur in non-quantitative variables, as when data are grouped by dichotomies, etc.

Since examples of these non-parametric methods are described in the paper by Moses, and in chapters of the Dixon and Massey text,[5] as well as in a more complete bibliography by I. R Savage,[6] the research student is referred to these sources for fuller detail on methodology than is presented here Our concern will be (1) to show the results of applying some of these methods to test data in two of the studies reported earlier in this book, and (2) to describe a new non-parametric method devised by Sol Haberman[7] of the University of Minnesota and designated as a method of partially ordered systems of expectations This method seems to have real promise in the analysis of empirical experimental designs of the kind reported in the present book.

Statistical analysis aims at isolating the factors which account for the variation in the measures obtained by systematic study For biometric data analysis of variance is employed, but since psycho-social measures often fail to meet the underlying assumptions of analysis of variance, we turn to non-parametric methods In these methods the foregoing difficulty can be overcome by arranging each set of values of the variable factors in order of size Thus, no assumption need be made as to the normality of the distribution of the original variate Friedman's χ^2_r can be applied to ranked data,[8] and may be employed in place of ordinary analysis of variance where

[4] John E Walsh, *J A S A*, September, 1952, pp 401–415, especially p 402

[5] W J Dixon and F J. Massey, *Introduction to Statistical Analysis,* McGraw-Hill, 1951, pp 247–263

[6] *J A S A*, December, 1953, pp 844–906

[7] Sol Haberman, "Special Distributions of Kendall's Tau," *Biometrika,* 1956

[8] M Friedman, "The Use of Ranks to Avoid the Assumptions of Normality Implicit in the Analysis of Variance," *J A S A*, December, 1937, pp 675–701 The data that are ranked may be standard deviations, as in Tables I and II, pp 676–677, or means, page 684.

there are two or more criteria of classification, i e., the data may be classified by kinds of treatment for different measures of effect. It is a far less arduous procedure than analysis of variance, since it requires less than one-fourth as much time,[9] but it is less efficient because it loses information by substituting ranks for numerical measures.[10] See Table 29, in which the following Friedman formula was used.

$$\chi^2_r = \frac{12}{np\ (p+1)}\ [\Sigma R^2] - 3n(p+1)$$

where $p =$ the number of Columns
$n =$ the number of Rows
$R =$ sum of column ranks

As Friedman says, " . . the method of ranks involves first ranking the data in each row of a two-way table and then testing to see whether the different *columns* of the resulting table of ranks can be supposed to have all come from the same universe This test is made by computing from the mean ranks for the several columns a statistic, χ^2_r which tends to be distributed according to the usual χ^2 distribution when the ranking is in fact random, i e. when the factor tested has no influence "[11] In the case of ties among the ranks the averages of the different values of the tied ranks are used[12] When there are only two columns, $p = 2$ (p in this procedure $=$ the number of columns in the table), the efficiency[13] of the method is only

[9] *Ibid*, p 696
[10] *Ibid*, pp 681–682
[11] *Ibid*, p 676
[12] *Ibid*, p 681
[13] Efficiency, to follow Dixon and Massey's concise definition, means "The estimates $\bar{x}$, s^2, and r, for the mean, variance, and correlation coefficient, respectively, are efficient in the sense that we obtain the maximum amount of information from the sample More precisely, these estimates have a smaller variance than any other estimates when our observations are from a normal population A statistic which has a variance smaller than any other estimate of the same population parameter is called an *efficient* estimate This is *statistical efficiency* The relative efficiency of another estimate can be measured by dividing its variance into the variance of the efficient estimate This quotient is called the efficiency E of the estimate The efficiency of all estimates other than the efficient estimates will then be less than 1 . "

63 7 per cent [14] In fact, Friedman's method is most effective for two or more criteria (classes) of analysis of variance; and for the case if $p = 2$, the method of Kruskal and Wallis [15] should be used In this latter procedure when the number of ranks is large (1) the maximum loss of information cannot be as great as 36 per cent; (2) the minimum loss of information is probably 9 per cent, and (3) the loss appears to decrease as the number of ranks increases

The Kruskal and Wallis [16] *H*-statistic is a non-parametric analysis of variance for one criterion of classification and is more appropriate for such data than Friedman's method. Like the usual procedure of analysis of variance, Friedman's non-parametric method summarizes a number of variables in one continuous operation but permits use of χ^2 tables to estimate probability, whereas the usual analysis of variance leads up to the *F*-ratio and the use of *F*-tables computed by Snedecor to determine probability values. These authors use the *H*-test on two of Snedecor's analysis of variance tables for one-criterion (class) analysis and obtain results close to the *F*-values obtained by Snedecor.[17] The only assumptions [18] underlying the use of the *H*-test are: (1) that the observations are all independent; (2) that those within a given sample come from the same population; and (3) that the *C* populations are approximately of the same form.

Since the Friedman method is a non-parametric analysis of variance, it yields an approximation to the results of the usual analysis of variance when the assumptions of the latter method have been

[14] Friedman, *op cit*, pp 682, 696, and W H Kruskal and W A Wallis, "The Use of Ranks in One-Criterion Analysis of Variance," *J A S A*, December, 1952, p. 602

[15] Kruskal and Wallis, *op cit*, pp 583–621

[16] *Ibid*, p 586

[17] *Ibid*, pp 588–589

[18] *Ibid*, p 585 It is worth noting in this context that in general, the assumptions underlying non-parametric methods are less numerous than those that underlie the analysis of variance or the *t*-statistic, and consequently the use of non-parametric methods are preferable for testing the results of empirical experimental designs In so far as these non-parametric methods yield conservative results, i e yield higher probability values of chance occurrence, they supply a "safety margin" against too easy inference But there is also the risk of making an error of the second kind—to accept the null hypothesis H_0, when in fact it is wrong, and therefore some other hypothesis such as H' is true.

met by the data. As Friedman points out, however, a non-parametric method which does not utilize all the information in the samples is less efficient than the conventional analysis of variance whenever the latter can be applied Nevertheless, the non-parametric method often yields probability results close to the probability values of F in the usual analysis of variance.[19]

If we now turn to Table 5, page 68, with the data of samples which depart considerably from the assumptions of the usual analysis of variance, and apply Friedman's χ^2_r to these measures of presumed multiple effects of improved housing, we obtain values of χ^2_r of the order of 4 2 for morale, social participation, and social status combined, and 6 6 for general adjustment, social participation, and social status combined, neither of which figures is statistically significant for 3 degrees of freedom Tables 29 and 30 show the detail. Explanations of technical adjustments appear in the footnotes to each table

TABLE 29 Changes in Measures of Effect (Including Morale) Associated with Public Housing [a]

Measures of effect	Treated and Non-treated Groups				
	Residents 1939	Residents 1940	Non-residents 1939	Non-residents 1940	$p = 4$
Morale [b]	60 1 (2)	60 2 (1)	58 0 (3)	56 6 (4)	$n = 3$
Social Participation	1 73 (1)	6 34 (4)	2 76 (2)	4 87 (3)	$\Sigma R^2 = 246$
Social Status	60 5 (1)	86 7 (4)	61 1 (2)	82 2 (3)	
ΣR (rank) [c]	4	9	7	10	
ΣR^2	16	81	49	100	

$$\chi^2_r = \frac{12}{np(p+1)}[\Sigma R^2] - 3n(p+1)$$

$$= 4\,2, \text{ for } 3\,d\,f,\ P_{30} > P > P_{20}$$

[a] Data from Table 5, p 68

[b] Rank order from the least desirable to the most desired size of score is shown in parentheses Morale is a reverse scale

[c] A 4 × 3 table is ordinarily a less satisfactory subject for χ^2_r analysis than one which has a larger number of cells

[19] Friedman, *op cit*, pp 684–686.

TABLE 30 Changes in Measures of Effect (Including General Adjustment) Associated with Public Housing[a]

Measures of effect	Treated and Non-treated Groups								
	Residents				Non-residents				
	1939		1940		1939		1940		$p = 4$
General Adjustment[b]	45	(1)	44 0	(2)	42 4	(3)	41 2	(4)	$n = 3$
Social Participation	1 73	(1)	6 34	(4)	2 76	(2)	4 87	(3)	$\Sigma R^2 = 258$
Social Status	60 5	(1)	86 7	(4)	61 1	(2)	82 2	(3)	
ΣR (ranks)	3		10		7		10		
ΣR^2	9		100		49		100		

$\chi^2_r = 6\ 6$, for 3 $d\ f$, $P_{.10} > P > P_{.05}$

[a] Data from Table 5, p 68

[b] Rank order from the least desirable to the most desired size of score is shown in parentheses General adjustment is a reverse scale

The chief points to be noted are (1) that the rank order numbers upon which χ^2_r is based reflect substantial differences in mean scores for social participation and social status, and by comparison, (2) for morale, the difference for residents is very small, i.e only 0.1 points on a score of 60. This would disappear entirely if the raw scores had been converted into standard scores[20] For general adjustment, the difference being 2 points or over, the standard scores would differ by 2 points, and for morale and general adjustment, the raw score differences, as compared to those of social participation and social status, are relatively insignificant. Nevertheless, in non-parametric analysis of variance all 12 scores in the tables are treated alike despite the existence of differences of magnitude of different orders (3) It will be noted that Table 30, for general adjustment, shows a χ^2_r which begins to turn toward statistical significance

The result is not statistically significant by this χ^2_r non-parametric method which, nevertheless, pools all the ranked measures in one series of operational steps Consequently, the answer is negative to

[20] E A Rundquist and R F Sletto, *Personality in the Depression,* University of Minnesota Press, 1936, p 389

the following question Are the different means of scores for different treatments at different dates by different score measures of response to treatment of such a nature that the observations could not have been drawn from the same universe whatever its distribution might be? Or, conversely, we may conclude that the observations support the hypothesis that for any one category of effect the numerical values of the mean scores on different treatments are, for all practical purposes, the same. Again, to put the matter in different words, the combined effect of column and row influences is such that the observed changes for the experimental group compared to the control group are of the same order of chance occurrence. This is also evident from the results of analysis by the *t*-statistic, since the multiple critical ratios on changes from 1939 to 1940 are 4.64 for non-residents and 5.16 for residents; or, to state the matter differently, the changes for non-residents and for residents are of an equally significant order and hence seem not to be associated with differential treatment This result corresponds in meaning with the result obtained by non-parametric analysis of variance, i e the *t*-analysis fails to support the conclusion that the measures were drawn from different universes.

Turning now to the summary Table 6, page 72, Friedman's non-parametric analysis of variance leads to a $\chi^2_r = 6.6$, which, for 3 degrees of freedom, is a result which is much more conservative than the multiple critical ratio of the *t*-statistics Thus we may accept the hypothesis that for any one category of effect the numerical values on different treatments are for all practical purposes the same

The explanation of why the χ^2_r statistic shows a lower order of statistical significance than the multiple critical ratio is (1) rank order ignores such relatively large differences as appear in the rows, whereas (2) the *t*-statistic, on which the multiple critical ratio is based, takes account of these differences Hence the *t*-statistic leads to higher degrees of significance. This method of analysis may be accepted provided one is willing to accept the argument by analogy on page 181 Furthermore, if we set up a table of the hypothetical order of magnitudes of means which accords with expectations that improved housing will show consistent gains on all measures of

TABLE 31 Summary Table of Changes in Measures of Presumed Social Effects of Public Housing [a]

Measures of effect	Treated and Non-treated Groups			
	Residents		Non-residents	
	1939	1940	1939	1940
Social Participation	1 73 (1)	6 34 (4)	2 76 (2)	4 87 (3)
Condition of Living Room	−0 20 (1)	3 00 (3)	3 50 (4)	2 20 (2)
Percentage Not Use-crowded [b]	50 00 (1)	94 00 (4)	55 30 (2)	70 10 (3)

$\chi^2_r = 6\,6$, for 3 $d\,f$, $P_{10} > P > P_{05}$

[a] Data from Table 6, p 72

[b] In Table 6, page 72, percentage use-crowded is shown In this table the complementary percentage is used so that the meaning of its ordering may be made consistent with the two other measures

effect (as in Table 32), then non-parametric analysis of variance would yield a $\chi^2_r = 10.4$ which, for 3 degrees of freedom, is $P_{02} > P > P_{01}$, and a closer approximation to the probability value obtained by use of the multiple critical ratio 4.97 than the $\chi^2_r = 6.6$ obtained from the observations.

TABLE 32 Data of Table 31 Arranged in Hypothetical Order as if the Expected Ordering Had Occurred

Measures of effect	Treated and Non-treated Groups			
	Residents		Non-residents	
	1939	1940	1939	1940
Social Participation	1	4	2	3
Condition of Living Room	1	4	2	3
Percentage Use-crowded	1	4	2	3

$\chi^2_r = 10\,4$, for 3 $d\,f$, $P_{02} > P > P_{01}$

In an earlier comment on the inconclusive small differences of such housing effect measures as morale and general adjustment, it

was pointed out that an interval of only one year from 1939 to 1940 may have been too short to reflect reliable changes in such complicated attributes as morale and general adjustment (see page 69, footnote 16). Let us then examine the data on 20 continuously resident families obtained in the follow-up study noted on page 174 over the years 1940 to 1942.[21]

When the original critical ratio or t-values of the differences on mean general adjustment scores are computed for 1940–1942, we obtain $t = 1.55$,[22] which is between the 20 per cent and the 10 per cent level, and hence not statistically significant. Let us see whether several other non-parametric tests applied to the same data for 1940–1942 are consistent with this result.

The Kruskal-Wallis non-parametric H-test for one criterion analysis of variance (on general adjustment measures only), yields $H = 2.01$, which has a probability value between the 20 per cent and the 10 per cent level for 2 degrees of freedom, and hence is not significant.

Next, the sign test[23] yielded an r-value of 6, which has a probability value of 25 per cent, and hence is not statistically significant.

Then the median test[24] was applied, and yielded a $\chi^2 = 1.6$, which, for 1 degree of freedom, has a probability between the 30 per cent and the 20 per cent level, and hence is not significant. In

[21] One family was misclassified as a resident in 1939, hence the correct total is 20 continuously resident families for the three dates 1939, 1940, and 1942.

[22] Since the difference between the means on general adjustment for the years 1939 to 1940 was slight, and larger for 1940 to 1942, this second period was chosen for test in the above calculations.

[23] Dixon and Massey, *op. cit.*, pp. 247–250, 324, and Moses, *op. cit.*, pp. 124–125, see also P. S. Olmstead and J. W. Tukey, "A Corner Test for Association," *The Annals of Mathematical Statistics*, 1947, pp. 495–518, A. Stuart, "The Power of Different Sign Tests," *J.A.S.A.*, September, 1952, pp. 416–424, and G. E. Noether, "Confidence Limits in the Non-Parametric Case," *J.A.S.A.*, September, 1949, pp. 342–355, and "Confidence Intervals for Median and Other Position Measures," *J.A.S.A.*, December, 1952, pp. 635–646.

[24] Moses, *op. cit.*, based on A. M. Mood's *Introduction to the Theory of Statistics*, McGraw-Hill, 1950, p. 394, see also two papers by John E. Walsh "Application of Some Significance Tests for the Median which Are Valid under Very General Conditions," *J.A.S.A.*, September, 1949, pp. 342–355, and "Confidence Intervals for Median and Other Position Measures," *J.A.S.A.*, December, 1952, pp. 635–646.

this test the median is computed from the combined observations of 1940 and 1942 Then a 2 x 2 contingency table is constructed with a number of cases with general adjustment scores above the median (+) and below the median (—) for each date; χ^2 is then computed from the table so constructed

TABLE 33 Follow-up on Twenty Families, 1940 to 1942

Family Number	General Adjustment			
	1940 Score	Rank [a]	1942 Score	Rank
1	35	9	45	24 0
2	55	36 5	44	22 5
3	50	31 5	44	22 5
4	64	39 0	56	38 0
5	29	3 5	40	16 5
6	46	26 0	36	10 5
7	51	33 5	43	20.5
8	51	33 5	39	14 5
9	34	8 0	37	12 5
10	66	40 0	42	18 5
11	49	29 5	37	12 5
12	54	35 0	55	36 5
13 [b]	50	31 5	47	28 0
15	42	18 5	32	6 0
16	40	16 5	46	26 0
17	46	26 0	39	14 5
18	49	29 5	43	20 5
19	33	7 0	36	10 5
20	31	5 0	25	1 0
21	29	3 5	27	2 0
Σ Scores	904		813	
Mean Score for $n = 20$	45 2		40 6	
ΣR's		462 5		357 5
R^2		213,906 25		127,806 25
R^2/n		10,695 31		6,390 31

[a] Rank in over-all array of 40 measures (2 × 20) thus, score 25 is #1, and score 66 is #40
[b] Case 14 omitted because ambiguous as between 1939, 1940, and 1942

Check on totals from Table 33 ΣR's = 462 5 + 357 5 = 820

from formula $\Sigma R = \frac{1}{2}N(N+1) = \frac{40}{2}(41) = 820$

$$H = \frac{12}{N(N+1)}\left[\Sigma\frac{R^2}{n}\right] - 3(N+1)$$

$= 2\ 01$, for 1 $d\ f$, $P_{20} > P > P_{10}$

where N = over-all total
R = rank

Correction factor for ties = 9988

$1 - \frac{\Sigma T}{(N^3 - N)}$, where ΣT = summation over all groups of ties = 72

$T = (t-1)t(t+1) = t^3 - t$

t = number of tied observations in the group

N = over-all total = 40

$$\frac{H}{1 - \frac{\Sigma T}{(N_3 - N)}} = \frac{2\ 01}{9988} = 2\ 012$$

Finally, Friedman's χ^2_r was computed for the data of Table 33, and yielded a value between the 30 per cent and 20 per cent probability values and hence was not statistically significant The conclusions from these different non-parametric methods when applied to the same data of general adjustment score changes between 1940 and 1942 are: (1) all four non-parametric methods confirm one another and are also in agreement with the original critical ratio or t-value of 1 5; and (2) the differences between scores in 1940 and in 1942 are not statistically significant, i e the differences are of an order that could occur frequently as chance fluctuations in random samples of the same size Although it may have little statistical significance in this one case, it is at least interesting to note that the estimate of probability by the t-test is nearer to the results of the H-test than to any of the other three non-parametric tests.

Sol Haberman's theory and technique of partially ordered systems of expectations is particularly adapted to testing empirical experimental designs in which the measures of effect are qualitative variables expressed as dichotomies, trichotomies, etc The structure of the design may be that of matched cases in control group comparisons, or "before" and "after" comparisons of measures on the same cases The hypothesis sets up systematic series of ordered expectations.

for example, in the problem of measuring the presumed effects of housing, the hypothesis is that improved general adjustment is associated with treatment, such that after a period of treatment, the measure of effect (scores on general adjustment scale) shows better adjustment of residents (lower scores, because in this example a reverse score scale was used) at the terminal date (end of treatment), as compared to the first date (before treatment began); and for direct score scales, such as social participation or social status, higher scores at the terminal date than at the beginning date, etc.

In using this method several points may be noted

(1) No assumption of a normal distribution of measures of effect is made; in other words, the method is distribution-free,

(2) The next point is to determine the probability of departures from the expected order of occurrence This determination is made by calculating the number of interchanges in position that would have to be made to bring the observed order into the order of expected events;

(3) Haberman has computed the probability distributions of chance occurrence of different numbers of such interchanges,

(4) This has been done for contingency tables showing two treatments (before and after, or experimental and control group) in column headings; and for row categories, 2, 3, 4, 5, and 6, sets of observations (data) In general, as the number of cells increases from 2 x 3 to 2 x 4, etc., the probability distributions tend towards normality [25]

The method has the following advantageous characteristics· (1) it is distribution-free; (2) it can be made free even of any assumption of random sampling when the situation so requires, as in paired non-random samples; [26] (3) it is a "goodness of fit" approach where the hypothesis or hypotheses to be tested are represented by a geometric

[25] See Appendix D

[26] The geometric models and tables of probability for these will be found in Sol Haberman's doctoral dissertation, *Special Distributions of Kendall's Tau* (τ), University of Minnesota, 1955, and in his article in *Biometrika,* 1956, by the same title These sources give his probability tables which may be used in testing small and non-random samples For random samples Haberman's table of probability is reproduced in Appendix D.

picture and an observed situation is tested for its consistency with the picture; and (4) it offers simplicity of computation in that interchanges in rank position are counted and their number judged as statistically significant or not from probability tables provided. Thus, less numerous assumptions are made than in some other non-parametric methods For all these reasons, therefore, Haberman's S-test is a very useful addition to other non-parametric methods in experimental designs, especially when applied in the free community situation. One assumption of the method is (independence), mutual exclusiveness among the contents of, the cells of the table;[27] another is that each of the possible rankings is equally probable.

Proof of his procedure appears in the references we have cited, but we may offer, in passing, a brief statement of the logic of his rationale. It begins with the familiar question· What are the ideal patterns or ideal constructs which are approached in observation and practice? The ideal patterns are the limiting patterns in form and number, but *not normative patterns* in the sociological sense. These limiting patterns are the frequencies of occurrence of different patterns (arrangements of things) *which are possible* The algebra of permutations and combinations gives us the form and numbers of these patterns which are possible. The units in these patterns are composed of combinations of specific dichotomies, trichotomies, etc. This use of the concept of limiting patterns invokes the theory of probability, so that probability as used in this context systematizes expectations to the principle of the limits of relative frequency.[28]

Application of the Haberman method to the data of Table 33 may

[27] In the Chi-square table, each cell (1) includes all *cases* of the class designated by its column and row heading (out of the total in the table), and (2) the contents of the cells are mutually exclusive Thus the test for independence is to determine whether or not a relationship exists between the vertical or column headings (classes) and the horizontal or row classes Obviously, if some of the cases that should have been in only one given cell of the cross-classification were placed in some other cell, then these data could not serve as a test of independence With respect to *measures* as entered in the cells of a table, independence means (1) the magnitudes of the measures in any cell should not depend upon the magnitudes of the measures in any other cell, and (2) when two or more scales are used there are no items or units which these scales have in common See Table 3, page 42, footnote *c*

[28] See chap VII, pp 183 to 186

be made as follows In the first place it will be seen that this table is a 2 x 20, or 40-cell table of before and after measures on the same cases Since the nearest applicable grouping in the Haberman table of Appendix D is a 2 x 4 table, or 8-cell table, we may divide the 20 cases of Table 33 into four random samples of five cases each (step 1) The four means of these four random sub-samples are then computed and placed in Table 34.

TABLE 34 Means of General Adjustment Scores [a]

Random Samples	1940		1942
1	43 8 a	⟶	39 8 A
2	49 2 b	⟶	41 8 B
3	37 6 c	⟶	36 6 C
4	50 2 d	⟶	44 4 D

[a] Arrows indicate the direction of expectations by hypothesis that $A < a$, $B < b$, etc , since the lower scores on this reverse score scale indicate improved adjustment If we had taken scores on social participation or social status, which are direct score scales, the larger magnitudes would appear in the second column with lettered capitals A, B, etc expressing the expectations

Next (step 2), we arrange the data by the order of size of means with the corresponding expectations in capital letters.

36 6 C
37 6 c
39 8 A
41 8 B
43 8 a
44 4 D
49.2 b
50.2 d

Since by order of scores the expectation (hypothesis) that $A < a$, and $B < b$, etc., is corroborated, no interchange of any item is needed to bring the observed order into the expected order, hence

$$S = 0$$

Now we turn (step 3) to Table A, Appendix D, and find that for a 2 x 4 table of 8 cells, that when $S = 0$, the corresponding probability is $\frac{2520}{40,320} = \frac{1}{16}$, or expressed as a ratio, it is 0625. Evidently the result of this analysis is a probability between the 10 per cent and the 5 per cent level, and hence is not statistically significant. It

will be remembered that the *H*-test yielded a result when applied to the same data that was not statistically significant. Incidentally, the *t*-test on the same means of Table 34 yields 2 56, which for 4 degrees of freedom is $10 > P > .05$, or the same result.

Let us now illustrate the procedure for a 2 x 5 cell table. To do this we may take the five cases of Sample 3 of Table 34, and form Table 35.

TABLE 35 General Adjustment Scores of Sample 3

Case Numbers	General Adjustment Scores 1940			1942	
18	49	a	⟶	43	A
5	29	b	⟶	40	B
20	31	c	⟶	25	C
19	33	d	⟶	36	D
17	46	e	⟶	39	E

25 C
29 b
31 c
33 d
36 D
39 E
40 B
43 A
46 e
49 a

Here only three of the sets of scores are in the expected direction, namely $A < a$, $C < c$, and $E < e$, while $D > d$, and $B > b$ To bring this observed order into accord with the expected order, we must move B up five positions, and move D up one position, hence

$$S = 6$$

From Table A in Appendix D, the probability of at most six interchanges for S is

$$\frac{113{,}400 + 113{,}400 + 151{,}320 + 188{,}880 + 225{,}000 + 256{,}320 + 281{,}040}{3{,}628{,}800}$$

Summing up, we obtain

$$\frac{1{,}329{,}360}{3{,}628{,}800} = \frac{1}{2\ 7297} = 3663,$$

hence the result is not statistically significant

Since empirical experimental designs for evaluating programs of social treatment in the free community situation are often limited to non-random groups, which may be the only kind available, it would be very useful to have a non-parametric method which could be used in such cases, and would at the same time provide a probability frame of reference to test the results. This has, in fact, been done by Haberman with appropriate modifications of the theory and technique which we have just described Haberman's *Biometrika* article provides the theory and geometric proof of such an extension, with corresponding probability tables for the *S*-values. We now turn, therefore, to examples to illustrate the application of the method to non-random groups.

To obtain a set of non-random sub-groups from the data of Table 33, these data are organized in an array from the case which shows the highest gain from 1940 to 1942 (this is Case 10, which changed from 66 in 1940 to 42 in 1942, an improvement of 24 points) to the case that shows the highest loss (this is Case 5, which changed from 29 in 1940 to 40 in 1942, a decline of 11 points) This array yields a non-random order. The 20 cases so ordered are then divided into four samples of five cases each and yield means of the small sub-groups as shown in Table 36. We now apply Haberman's method for non-random samples to the data of Table 36, noting that in this example the expected orders of score magnitudes are required to meet conditions *additional* to those required in Tables 34 and 35.

We shall test the composite hypothesis that not only is there an increase in general adjustment between years but also there is a rank order of adjustment among the samples within each year For this reason we must pretend, in our illustration testing this hypothesis, that the samples were not selected directly on the basis of increases in general adjustment scores but on some criterion or set of criteria which we suspect correlate highly with the measure of increases The composite hypothesis, then, consists of two parts

Part 1 System of expectations used on random small groups.

A precedes a, (written as A , a , etc in Table 36) [29]

[29] For direct score scales in which greater magnitudes signify gains, these orders would be reversed to express the expectations

B precedes b,
C " c,
D " d,

Part 2 System of expectations *added* when non-random samples are to be tested, as in Tables 36, 37, and 38.
A precedes B precedes C precedes D (written A, B, etc.)[20]
a " b " c " d (written a, b, etc)

Keeping in mind the composite hypothesis we may now proceed to a test of the data of Table 36 by means of the following systematic steps·

(1) Over-all order of observed mean scores from both columns	(2) Expected order by hypothesis	(3) Number of interchanges to satisfy (2)	(4) Final order after interchanges by (3)
34 2, a	A, a	3	A
38 8, D	B, b	0	a
39.6, B	C, c	0	B
40 8, A	D, d	0	b
42.8, b	A, B	0	C
43 4, C	B, C	0	D
51.2, c	C, D	3	c
52.6, d	a, b	0	d
	b, c	0	
	c, d	0	
	Total	$= S = 6$	

It would take three interchanges to bring A above a, and an additional three to place D below C. This total of six interchanges is the minimum number needed to convert this observed ranking into a ranking which is consistent with expectations The ranking we achieve by these manipulations appears in the last column above.

Haberman's table in *Biometrika* yields a probability of getting, at most, an $S = 6$, as follows

$$\frac{14 + 56 + 150 + 318 + 587 + 963 + 1448}{40{,}320} = \frac{3{,}536}{40{,}320} = 0877,$$

TABLE 36 Means of General Adjustment Scores Four Non-random Groups

Samples of Small Groups from Array [a]	General Adjustment Scores (means) (1) 1940	(2) 1942
	— [b]	40 8 , A [c]
Lowest Five Cases	34 2 , a	39 6 , B
Lower Middle	42 8 , b	43 4 , C
Upper Middle	51 2 , c	38 8 , D
Highest Five Cases	52 6 , d	— [b]

[a] Of 20 cases of Table 33, by magnitude of observed improvement 1940 to 1942 in general adjustment

[b] Note the non-parallel arrangement of column entries, so that the entries are on diagonals in the direction of expectations This is an essential part of the technique as explained in the article in *Biometrika*

[c] Not only is the 1942 score expected to be less than the 1940 score for each sample in this table, but the scores are expected to fall in an order within each year (column) from smallest to largest score

which is not statistically significant

Let us now try out the method on two other non-random sub-groups of Table 33, taking this time two small sub-groups of four cases each. Data are shown in Tables 37 and 38.

This time we shall assume that in Table 37 we have good reason to expect that Case 17 precedes 18, which in turn precedes 3, which in turn precedes 20 for both 1940 and 1942 Note that this assumption happens to follow *approximately* the observed order of these cases in the array This is mere coincidence in the present example; it is noted here because it will be seen that the observed orders of scores in the first column of Tables 36 and 38 are similar, i e, they happen to fall in *regular sequence* from smallest to largest.

TABLE 37 General Adjustment Scores of Sample at Mid-range

Case Numbers	General Adjustment Scores (1) 1940	(2) 1942
	—	39 , A
17	46 , a	43 , B
18	49 , b	44 , C
3	50 , c	25 , D
20	31 , d	—

Since the steps of procedure to test the data of such a table were shown in detail for Table 36, we shall abbreviate here and merely indicate the main points

Over-all order of observed scores	Ranking achieved by interchanges
25 , D	A
31 , d	B
39 , A	C
43 , B	D
44 , C	a
46 , a	b
49 , b	c
50 , c	d

It takes six interchanges to bring d down below c, after which it then takes an additional three to put D below C This total of nine interchanges is the minimum number needed to convert the observed ranking of Table 37 into a ranking which is consistent with expectations.

The probability of getting at the most, $S = 9$, is calculated as

$$\frac{14 + 56 + 150 + \quad + 2599 + 3151}{40{,}320} = \frac{11{,}295}{40{,}320} = 2802,$$

which is not statistically significant.

In Table 38 we shall assume that we have good reason to expect that for both 1940 and 1942 Case 11 precedes Case 8, which precedes Case 2, which in turn precedes Case 10. This is, of course, just Part 2 of the composite hypothesis testing the arrangement of the table The *S*-tests of two non-random samples of Tables 36 and 37 indicate that the values in the table when ranked are not to be taken as consistent with expectations However, the ranked values pertaining to the third non-random sample entered in Table 38 give the following result

Over-all order of observed scores

37 , A
39 , B

If we interchange C and D, the ranking is consistent with the expectations Referring to the table, the

42 , D
44 , C
49 , a
51 , b
55 , c
66 , d

probability of getting at most one interchange is $\frac{14+56}{40,320} = \frac{1}{576} = 0017$, which is statistically significant at any of the conventional levels used in practice.

TABLE 38 General Adjustment Scores of Four Highest Improvement Cases

Case Numbers	General Adjustment Scores (1) 1940	(2) 1942
	—	37 , A
11	49 , a	39 , B
8	51 , b	44 , C
2	55 , c	42 , D
10	66 , d	—

The foregoing applications of Haberman's method to non-random groups relates to general adjustment measures on a *reverse* scale. But another scale was also used as a measure of presumed effects of good housing—a measure of change in the condition of the furnishings of the living room of the homes When scores on this *direct* scale are obtained for 1940 as compared to 1942 and the significance tested by the method of Table 36, the A, B, C and D values are expected to be of greater magnitude than the a, b, c and d values Proceeding in this manner we obtain $S = 6$, or at most six interchanges, to bring the observed order into correspondence with the expected order. Thus for $S = 6$, the probability is

$$\frac{14 + \ldots + 1448}{40,320} = \frac{1}{11\,4} = 0877$$

and the result is not significant

One final application if the reader will now turn to Table 5, page 68, wherein the difference between social status scores of the 44 resident families (column 4) for 1939 to 1940 are shown, and apply the S-test, we obtain $S = 2$, with a probability of

$$\frac{14 + 56 + 150}{40{,}320} = \frac{1}{183\ 2} \text{ or } 0054.$$

This compares with a probability of .001 for the critical ratio of 4 27, shown in that table The S-value of two interchanges was obtained by the method of Table 36, so as to apply Haberman's 2 x 4 table probability model to the 2 x 44 table of data. The 44 cases were subdivided into 4 groups of 11 cases each. Then the mean of each subgroup was computed and the means arranged in a 2 x 4 table like Table 36 From this point on the steps taken in Table 36 were followed and the S-value of 2 was obtained. These two illustrations indicate that significant probability values may be obtained by this method when the data so indicate.

In conclusion, suppose that in some given study we obtain a like result, and we know that the sample of four cases is non-randomly chosen from some population where on the basis of the non-randomness we can set up certain expectations in addition to the other expectations to be tested, can we generalize the results to the larger population on the basis of this evidence? Certainly not!

But one thing we can state with certainty based on this evidence: Our results may be considered as identical with an expected pattern of events since there is a negligible chance departure from that pattern

In bringing this chapter to a close, and having in mind the results obtained in the example of Table 38, we end our comment on the same note as that which ended Chapter VIII, except that now we have a method which has eliminated one of the double analogy handicaps which stood in the way of determining probability The Haberman tables for non-random samples or small groups, by introducing testable expectations through the non-randomness, supply a theoretical model for direct comparison with empirical results.

Chapter X

THE EX POST FACTO DESIGN: REPLICATION AND EXTENSION

THE exploration of antecedent-consequent relationships or *probable causation* is facilitated by use of the *ex post facto*[1] experimental design. In the present study it was assumed that rentals may be taken as a rough index of a bundle of social and economic factors causally related to crowding in dwelling units. There is empirical evidence from previous studies of the relationship of crowding to certain social and structural factors, including rental rates. For example, Schmid[2] found a correlation of $r = -.555$ between rentals and percentage of crowded dwelling units in the Twin Cities in 1934 by enumeration districts of the 1930 census; other relationships were found by him and several other investigators.[3] In the present study a correlation of $r = -.68$ (corresponding to Schmid's result) was found between rentals and per cent crowded for the Borough of Manhattan, New York City, in 1934.

These empirical results suggest the hypothesis: there is an inverse relationship (correlation) between rentals and crowding. This measure of relationship correlates rentals and crowding *for a single date* and *does not* measure antecedent-consequent relationship. For the Borough of Manhattan, using census tracts as units of area, a correlation of $r = -.31$ was obtained between *changes* in rentals and associated *changes* in percentage crowded[4] from 1934 to 1940. This

[1] F. Stuart Chapin, Clarence A. Johanson, and Arthur L. Johnson, "Rental Rates and Crowding in Dwelling Units in Manhattan," *A.S.R.*, February, 1950, pp. 95–97; see also chapter v, especially pp. 95, 124–138. The present study is an extension of a previous exploratory study of all boroughs of New York City made by Elizabeth Y. Arneson, Raymond O. Farden, Clarence A. Johanson, Arthur L. Johnson, and Swan Ling under the direction of F. Stuart Chapin.

[2] Schmid, *Social Saga of Two Cities, op. cit.*, p. 253.

[3] See chapter v, pp. 127–128.

[4] Dwelling units having 1.01 or more persons per room were defined as crowded.

correlation, based upon changes between two dates, approaches more closely to an analysis of possible cause and effect relationships, but of course, it offers no proof of cause and effect Since the abuse of correlation coefficients has sometimes taken the form of accepting a correlation as equivalent to a measure of causation, it may be interesting to subject the data of rentals and crowding on two dates for the same areas to a method of analysis with greater capabilities for the discovery of assumed cause and effect relationships. This method is the method of *ex post facto* experimental design, in which some present effect (i e, crowding in 1940) is traced backward to an assumed causal complex of factors or forces (i e, rentals) at a prior date (1934), using for this purpose such records as are available (i e., census tract data of Manhattan), since no new measures of the past can be made in the present, and relying upon procedures of selective control (i e, matching the experimental group against the control group on per cent home ownership in census tracts).[5]

An examination of Table 39 shows results that support the hypothesis. This table demonstrates antecedent-consequent relationships more clearly and more simply than the correlation of $r =$ $-.31$, between *changes* in rentals and *changes* in crowding for the entire Borough of Manhattan. It shows that in an experimental group the low rentals ($19 93 in 1934) and high crowding rate (31 33 per cent in 1934) were followed by lower crowding (24 59 per cent in 1940) with slightly higher rentals ($22 95 in 1940); and the lowest crowding (7.47 per cent in 1934) associated with highest rentals ($93 73 in 1934) were followed by higher crowding (15 24 per cent in 1940) and diminished rentals ($85 33 in 1940). Not only does this experimental design show the inverse relationship between rentals and crowding more clearly and more simply than does the negative correlation of — 31, between changes, but it permits a test of the significance of the changes in the two associated variables. A decline in rentals of $8 40 for the period 1934 to 1940 has a critical ratio[6] of 4 54, highly significant, and the corresponding increase in crowding

[5] See chapter II, p 33

[6] The critical ratios are corrected for correlation between the variables attributable to matching on the per cent home ownership

TABLE 39

Groups Matched on Home Ownership	1934 (1) Areas of Extreme Low Crowding with High Rentals $N = 18$		1934 (2) Homogeneous Rent Areas $N = 35$		1940 (3) Areas of Extreme Low Crowding with High Rentals $N = 18$		1940 (4) Homogeneous Rent Areas $N = 35$		Changes (5) Areas of Extreme Low Crowding with High Rentals $N = 18$		Changes (6) Homogeneous Rent Areas $N = 35$	
E—gr												
Crowding (1)	7 47%		31 33%		15 24%		24 59%		+ 7 77%	(8 82) [a]	− 6 7%	(−7 39)
Rentals (2)	$93 73		$19 93		$85 33		$22 95		−$ 8 40	(4 54)	+$3 02	(4 96)
C—gr												
Crowding (3)	6 13%		15 43%		6 74%		17 33%		+ 0 61%	(0 84)	+ 1 90%	(+1 9)
Rentals (4)	$103 08		$46 50		$152 83		$46 29		+$49 75	(4 60)	−$0 21	(−0 22)
Diffs												
Crowding (5)	1 34%	(1 13) [a]	15 9 %	(8 85)	8 50%	(8 53)	+ 7 26	(+4 1)	+ 7 16%	(6 28)	− 8 64+	(−6 32)
Rentals (6)	−$9 35	(−1 70)	−$26 57	(−35 37)	−$67 5	(−4 8)	−$23 34	(−18 04)	−$58 15	(−5 3)	+$3 23	(2 85)

[a] *CR*'s shown in parentheses and all corrected for *r*'s introduced by matching It will be observed that the dependent variable crowding is approximately equated between experimental and control groups for 1934, since the difference of 1 34 per cent [col (1)] is statistically not significant (critical ratio, 1 13) Also the independent variable rental, shows for 1934 [col (1)] no statistically significant difference between means In column (2) however, the differences between means of both dependent and independent variables for 1934 are statistically significant This situation results from the nature of the category "homogeneous rent areas " Homogeneity was determined from a study of the incidence of tracts in five dollar rent classes on a map of Manhattan showing the ecological homogeneity of the tracts in four rental classes For this reason the differences in 1934 of column (2) could not be equated but were taken as they occurred

of 7 77% has a highly significant critical ratio of 8 82! Meanwhile, the associated changes in the control group were indeterminate

The convergence of evidence supports, rather than proves, the hypothesis of an antecedent-consequent relationship of an inverse type, i.e , low rentals at an earlier date are followed by a high percentage of crowding at a later date, and vice versa This evidence consists of (1) consistency with the magnitudes of correlations made in other studies, (2) consistency in the direction of relationship—all are inverse correlations; (3) critical ratios of change are beyond ordinary chance expectation, and (4) these inverse relationships hold for more homogeneous sub-groups, i.e , after one related factor is controlled by matching.

The direction of changes in rentals and in corresponding crowding was made for two samples analyzed The experimental group of columns 1, 3 and 5, rows 1 and 2 in the table, was selected from the extremes of increasing crowding and decreasing crowding, with the corresponding rentals, and the experimental group of columns 2, 4 and 6, rows 3 and 4, was selected to represent low rent areas as contrasted with middle rent areas 1934–1940 and their corresponding crowding rates. In both experimental groups the inverse relationship between rental and crowding is clearly shown, together with the statistical significance of both the *differences* between experimental and control groups at each date, and the significance of the *changes* between dates for each group

Whether a statistically significant sequential relationship between the variables indicates a causal relationship or a mere chance relationship, cannot be established (1) by the statistical data in themselves alone, or (2) by only one exploratory study such as the present investigation

Fundamental to proof of casuality and the power to predict results are· (a) introduction into our explanation of some conceptual scheme or theoretical considerations which entail the use of relevant data and experience not part of the immediate observations, and (b) repetition of similar studies with like results The theoretical considerations begin with the *assumption* that rentals are taken as a rough index of a bundle of socio-economic factors which *may* operate as a

complex pattern of causal factors, and then to utilize relevant data of experience, such as the known association of low rentals with low income, poor housing expressed in sub-standard sanitation, need of structural repairs, age and decreptiude of structure, etc Furthermore, as Dr. Knupfer has said ". . it appears that 'approximate rent' is the best index of 'general status,'"[7] a conclusion derived from factor analysis of 15 factors of status, including ammains, automobile, education, income, occupational, telephone, etc Thus, all that can be claimed for the present exploratory study are the following (a) the utility of the *ex post facto* experimental design as a method for clarifying and simplifying some *assumed* causal relationship, and (b) the verification or disproof of an hypothesis in a *particular example* in a *given place* for the *stated period* of the investigation

Critics of empirical social research have often asked that the results be tested by repeating the investigation[8] The present study offers one example of confirmation of results previously published, by a new study using a similar design under like conditions. Essential to additive research or cumulative investigations is to test the results of one study by (1) using the same method, on (2) like data, (3) investigated under similar conditions, and (4) for the same period In meeting these conditions the present study is a contribution to additive research.

In the preceding pages appear the results of an ex post facto experimental design which showed that a decline in rental rates 1934–1940, in the Borough of Manhattan, New York City, was antecedent (1934) to a statistically significant increase in crowded dwelling units (1940), and vice versa. In the present study a similar design was used to test this inverse relationship in time based on the same kind of data but for the Twin Cities (Minneapolis and St. Paul, Minnesota), 1934–1940

Table 40 shows the similarity of association between rentals and

[7] Genevieve Knupfer, *Indices of Socio-Economic Status A Study of Some Problems of Measurement,* New York, 1946 (Ph D Dissertation, Columbia University), pp. 104–107, 194

[8] F Stuart Chapin and Sheldon Stryker, "Confirmation of Results of an Ex Post Facto Experimental Design by Replication," *A S R*, October, 1950, pp 670–672

crowding for two urban areas when measured by correlation coefficients.

In view of the different geographic locations and the differences in size of urban areas compared, the correspondence between the two results suggests that the formulation of a general law of social relationship in the ecology of urban areas in the United States has been approached. This evidence of correlation suggests a point made in the first study (p. 95): while correlation based on *changes* between two dates approaches a description of causal relationship, it does not demonstrate as clearly the sequential relationship between

TABLE 40. Correlations

	Between rentals and crowding, 1934	Between changes in rentals and crowding, 1934–1940
Borough of Manhattan	−.68	−.31
Twin Cities	−.65	−.34

factors as does the ex post facto design in which antecedent-consequent conditions are made visible as in Table 41.

An examination of Table 41 shows that the results of the replication tend to confirm the original findings. When low rentals are associated with high degrees of crowding (experimental groups), a statistically significant change in these rentals is followed by an inverse change in crowding, which is statistically significant (rows 1 and 2); but when similar initial degrees of crowding (the *t*-values of differences not significant for 1934) are associated with *increases* in rentals (the control groups), the change in crowding is not statistically significant. Table 42 segregates the essential data of changes from Table 41 for more specific analysis.

Replication of the original experimental design thus supports the following conclusions drawn from a comparison of these two studies: (1) confirmation of the results in all matters of measured antecedent-consequent relationship; (2) additional evidence is pre-

sented to illustrate the efficacy of ex post facto experimental design as a research procedure which clarifies an antecedent-consequent relationship, and (3) support is given to the belief that the formulation of a general law of social relationship in the ecology of urban areas has been made, viz., (+) (−) changes in mean rentals in urban areas, as an index of a bundle of socio-economic factors, are followed by (−) (+) changes in mean per cent crowding; or that there is an inverse relationship (causal?) in time sequence between rentals and crowding in dwelling units in these urban areas

Two types of units have been used in the ex post facto studies described in this book (1) ecological groups of human individuals who are enumerated as residents at some given date in census tracts or census enumeration districts, or in some similar urban area-unit of ecological significance This was the case of the studies described earlier in this chapter, and in Chapter V, pages 124 to 139 Studies based on this type of unit present methodological difficulties that are not easy to overcome, although W S Robinson has attempted to make a transition from the data of ecological correlations to corresponding correlations among the unknown individuals who were the residents of the area units [9] (2) The human individual as the unit, followed through some part of his personal history This was the case of the two studies described in Chapter V, pages 95 to 124; and is also the case of the studies by Malcolm B Stinson and David O Moberg to be described presently.

The usual longitudinal studies [10] in medical and psychiatric research, and projected experimental designs with human subjects by research sociologists, allow us to observe that which happens as it takes place, and permit manipulative control by the observer of successive (present) conditions as causal factors, or as factors antecedent to some future effect In such studies we have to wait for the effects. During the run of such studies cases are lost from such

[9] In "Ecological Correlations and the Behavior of Individuals," Robinson has devised a conversion formula to estimate from ecological correlations the corresponding correlations among individuals This technique was not available when the data of the foregoing studies were analyzed See *A S R*, June, 1950, pp 351–357

[10] Neva R Deardorff, "Longitudinal Studies of H I P Experience, Background and Some Findings of Pilot Study," *J A S A*, June, 1954, p 346

natural causes of attenuation as illness, death, mobility, and refusals to give information In studies of human relations by comparisons of before and after measurements on an experimental and a control group, the losses of cases from natural attenuation factors, such as death, refusals of information, illnesses, and mobility, are so large as to destroy any initial randomization Whenever such studies on human populations have attained a higher degree of objectivity, quantification, or measurement, it may be possible to utilize such techniques as described by C R. Rao in his "General Methods of Analysis for Incomplete Block Designs."[11]

In the ex post facto approach we start out from two present groups of persons who show contrasting situations, and try to trace backward in time the factors which it is assumed may have operated to produce the present contrasting conditions Adults, or aged persons in particular, taken at a given date are the survivors of long periods of antecedent influences, since selective factors have taken their toll and there are no losses of present cases to occur. Each survivor has his own personal history of experience over a long past The records of these antecedent forces appear in official marriage and birth records, records of past illnesses, records of educational achievement, occupational history, and in such past events as deaths of family members, in memberships and offices held in various forms of social organizations, etc. Historians and political scientists have used such records to write biographies of unusual people, and physicians, teachers, and social workers have collected and recorded factual data on personal histories of the masses. Inferences are made concerning the influence of intangible factors, which sometimes help explain present personality variations and individual success or failure in a given person's adjustment to social conditions and to economic opportunities

It is my conviction, however, that a greater degree of objectivity may be attained by a more systematic probing of the past which is possible by ex post facto design of study In these studies the empirical ordering of steps of analysis may be guided by some frame of reference other than the literary, intuitive, qualitative, or historical,

[11] *J A S A*, December, 1947, pp 541–561

TABLE 41 Comparison of First Design of Original Study and First Replication Design

(1)	1934 (2) Original	1934 (3) Replication	1940 (4) Original	1940 (5) Replication	Changes[b] (6) Original	Changes[b] (7) Replication
Groups Matched on Home Ownership	Areas of Extreme Low Crowding with High Rentals[a] $N = 18$	Area of Low Crowding[b] $N = 10$	Areas of Extreme Low Crowding with High Rentals $N = 18$	Areas of Low Crowding $N = 10$	Areas of Extreme Low Crowding $N = 18$	Areas of Low Crowding $N = 10$
Rows						
E—gr Crowding 1	7 47%	9 60%	15 24%	6 86%	7 77% (8 82)[c]	−2 74% t = 3 10 01 > p > 001
Rentals 2	$ 93 73	$30 41	$ 85 33	$35 63	−$8 40 (4 54)	+$5 22 t = 2 25 05 > p > 02
C—gr Crowding 3	6 13%	8 55%	6 74%	8 77%	61% (0 84)	+ 22% t = 159 9 > p > 8
Rentals 4	$103 08	$28 86	$152 83	$31 69	$49 75 (4 60)	+2 83 t = 1 43 2 > p > 1
Differences[d] Crowding 5	1 34% (1 13)[e]	1 05% t = 805 5 > p > 4	8 50% (8 33)[e]	−1 91% t = 1 71 2 > p > 1	7 16% (6 28)	−2 96% (7 08)
Rentals 6	−$9 35 (−1 70)	$1 55 t = 539 7 > p > 6	−$67.50 (−4 8)	$3 94 t = 1 48 2 > p > 1	−$58 15 (5 3)	+$2 39 (2 97)

[a] In the original study of Manhattan, the experimental and control groups consisted of samples of census tracts from areas of extreme *low* crowding associated with *high* rentals Since census tracts were available for the Twin Cities only in 1940, comparable areas for 1934 were obtained by adjusting 1930 enumeration districts to equivalent census tract areas for 1934

[b] In the replication study of the Twin Cities, the experimental and control groups consisted of samples of census tracts from areas of low crowding in 1934

[c] Figures in parentheses in columns 2, 4, and 6, are critical ratios, in columns 3, 5 and 7, these are t-values

[d] For 1934 it will be noted that differences between groups are *not* statistically significant, hence the groups begin the experimental test period 1934–1940 on an equal basis in rentals and crowding

[e] For 1940, differences between groups *are* statistically significant (for Manhattan) or approach statistical significance (Twin Cities)

[f] Changes and differences in changes 1934–1940 *are* statistically significant for both experimental groups, meanwhile, changes in control groups are *not* statistically significant (except for rentals in Manhattan—see footnote *b* Table 42 for explanation)

Table 42 Changes in Rentals and Crowding, 1934–1940 [a]

(1) Experimental Groups	(2) Original Changes	Cr ratios	(3) Replication Changes	t-values	(4) Difference Between Changes Original	Replication
1 Crowding	+ 7 77%	+8 82	− 2 74	−3 10		
2 Rentals	−$ 8 50	−4 54	+$5 22	+2 25		
					crowding 7.16 (6 68)	2 96 (7 08)
					rentals $58 55 (5 30)	$2 26 (2 97)
Control Groups						
1 Crowding	+ 0 62%	+0 84	+ 0 22%	+1 59		
2 Rentals	+$49 75	+4 60 [b]	+$2 83	+1 43		

[a] In all comparisons the groups are matched on per cent home ownership of census tracts (+) indicates an increase in rates or per cents, and (−) indicates a decrease

[b] This change of $49 75 in rentals from $103 08 in 1934 to $152 83 in 1940 for Manhattan is the only statistic which fails to correspond to the trend noted in the other 7 statistics *This significant change in rentals with no corresponding change in diminished crowding* (+0 84), may be explained in this manner when rentals are high in the first date (1934 is $103 08) and still higher at the second date (1940 is $152 83), without an associated decline in crowding, this may indicate that there is a critical rental rate ($93 73, or 10 per cent lower than the $103 08 of Manhattan for 1934) from which an increase or decrease may be expected to follow in crowding, so that higher rentals both before and after tend to be followed by no appreciable change in crowding, since the critical level has been passed

in the conventional sense of these terms Such a frame of reference looks to the possibility of discovering some mathematical model which would then provide the logical frame of reference and the rationale for application of methods of statistical analysis so that it might be possible to isolate specific factors or patterns of factors operating in the past, but related to the differential conditions observed in the present If some mathematical model is to be found, the data of ex post facto studies must be quantified, expressed as variables or as proportions, and if the relevant factors are clearly qualitative, they should be classified at least into dichotomies, trichotomies, etc

Ex post facto design, then, utilizes a comparative method of analysis by defining a so-called "experimental" group which has been exposed to some prior treatment, or to some prior conditions or factors, to be set along side of a "control" group, not exposed to these antecedent factors, or exposed in a differential manner to these antecedent factors Such a "control" group [12] serves the purpose of a reference group for possible measurement of effects and probable inference as to the operation of factors in the past. When randomization of individuals into such groups is not possible, and in this design randomization seldom is possible, resort is made to matching on independent variables or factors, to obtain some degree of control of such factors, and to increase the degree of homogeneity If probability tests are applied to obtain a measure of statistical significance of the observed past differences between such groups, there still remains the problem of logical validity of making statistical inferences from non-random samples, or from groups in which the data are found, rather than from classes into which the cases have been assigned at random In all such studies application of probability is purely analogical, as I have contended elsewhere [13]

[12] In this context the "control" group can seldom be a wholly "unexposed" or "untreated" group, but it still serves the purpose of a reference group in the comparisons which the nature of the data permit

[13] *J A S A*, September, 1949, pp 460–462, "A Clarification of Experimental Designs in Sociological Research," *Sociology and Social Research*, January–February, 1950, pp 159–168, "Experimental Design in Sociology," *Social Forces*, October, 1950, pp 25–28, "The Experimental Method in the Study of Human Relations," *Scientific Monthly*, February, 1949, pp 132–139

As a matter of fact these ex post facto designs might stand to gain materially from developments in the area of non-parametric statistical method It will be my effort now to describe briefly two later studies, pointing out their limitations, but also endeavoring to point up the possibilities of greater efficiency by use of "simulated" experimental design.

Medical Care Costs Among the Aged

The first is a study of high to low cost of medical care following upon long personal histories of prior events, conditions, or factors among 354 aged persons 65 years of age and over, chosen as a stratified random sample from 7100 aged persons who received Old Age Assistance (OAA) including medical care in Hennepin County, Minnesota, for the twelve-month period ending in October, 1950. From an array on costs of medical care, comparisons were made of halves, upper and lower quartiles, and upper and lower deciles to ascertain whether the percentage incidence of specific prior medical diagnoses and certain social factors of past personal history differed for the two groups This study was a doctoral dissertation in social work at the University of Minnesota, 1951, by Malcolm B Stinson.[14]

Table 43 shows the variations in t-values between halves (median cost of medical care for 12 months ending November 1950 was

[14] Malcolm B Stinson, *Prediction of Medical Costs for the Aged,* Ph D Thesis, December, 1951, University of Minnesota, pp 109

OAA is financed by combined Federal-state-county funds The person must be 65 years of age or older, with a specified length of residence in the state and county, and in need of assistance "Need is established through the so-called budget deficit process, which means that the applicant's requirements for food, clothing, shelter, medical care, incidentals, and the like are calculated, and from this total the amount of his monthly income and resources is deducted The balance of requirements over income establishes need" The maximum payment was $55 for maintenance Payments for medical care were in addition to maintenance but with no ceiling on the amount (pp 2–3)

The 1950 total spent from state and county funds was at the rate of $1,893,005 per year (p 4) Individual payments for the twelve-month period ranged from $1 to over $2,000, with high concentration at low costs, and very small numbers receiving very large amounts (p 6) The average per individual of these medical payments was over $242 per month, and the median was $72

TABLE 43 Summary of Factors Which Discriminate Between High and Low Cost Cases for Medical Care among Recipients of Old Age Assistance in Hennepin County, 1950, Showing Critical Ratios[a]

Factors	Critical ratio of differences between high and low cost cases in[b] Halves	Quartiles	Deciles
1 Marital status	(N = 177)	(N = 91)	(N = 35)
a Single	x	1 67	x
b Married	x	−2 54	−1 75
c Widowed			
d Separated or divorced	−2 0	x	x
2 Children	x	x	x
a With	x	x	x
b None	x	2 27	x
3 Children paid bills	x	x	2 63
4 Principal diagnosis			
a Heart diseases	−1 67	−1 27	−2 88
b Arthritis and fractures	3 33	3 03	2 60
c Diseases of brain and nervous system	2 00	3 26	2 91
d Cancer—all types	2 27	2 12	All
e Urogenital, except cancer	0 0	2 33	All
f Diabetes Mellitus	1 90	1 85	All
g Asthma and bronchitis	−2 00	−All	−All
h Nutritional, including anemia	−0 50	−0 55	−All
i Teeth and glasses only	−1 62	−4 56	−All
j Cataracts	0 0	−1 30	0 0
k All other	−1 25	−2 12	−All
5 Minneapolis General Hospital never used it	−2 11	x	x
6 Usual lifetime occupation Craftsmen and operatives	x	x	2 33

[a] Data selected from Table 35, pp 96–97, Malcolm B Stinson, *Prediction of Medical Care Costs for the Aged A Problem in Social Work Administration*, a Ph D Thesis, University of Minnesota, December 1951 Differences not statistically significant indicated by x Critical ratios (CR's) are *t*-values It will be noted that section 4—Principal Diagnosis—shows for diseases b, c, d and g, persistent differences with statistical significance for all groupings of data, and a, f, i and k, show statistical significance for some groupings only

[b] A positive CR discriminates in the direction of the high cost cases A negative CR discriminates in the direction of the low cost cases

$72 00), quartiles, and deciles. For social factors in past personal history, no statistically significant differences at halves were obtained, but significant differences were found at quartiles for aged persons with no children (upper), who had been married (quartiles); or divorced or separated (halves); and at deciles where children had paid the bills (upper), or where the usual lifetime occupation had been that of industrial operatives or craftsmen (upper decile). For the category of principal medical diagnosis, statistically significant differences in percentage incidence were found for arthritis and fracture, cancer of all types, diseases of the brain and nervous system, asthma and bronchitis, in that order, followed by statistically significant differences at quartiles and deciles. For other diagnoses, the trend was irregular; not significant at halves, but in some cases significant at quartiles or at deciles It does appear, however, that for these aged persons, statistical significance found at halves is usually increased in magnitude at quartiles and deciles. This group of aged persons living as of November 1950, although a random sample of a population of survivors, cannot logically be regarded as also a random sample of survivors at earlier periods of 5, 10, 15, etc, years before We have no data on the selective factors which operated at those specific previous dates, and this lack of data complicates the problem of inference and places severe limitations upon any generalizations we might wish to formulate. In a rough manner, however, the data do seem to point to the possibility of prediction of high or low medical costs of any present group of aged survivors from antecedent independent factors of principal medical diagnosis and selected social factors

From the first example, it seems logical to assume that the greater the difference between the means or proportions on some measure of assumed present effect, the greater the probability of finding a corresponding statistically significant difference between the means or proportions on some independent variable or factor of past experience In the first example, the measure of effect was the cost of medical care expressed in dollar values, but in the example that follows, a psychometric measure of personal adjustment was used to describe the dependent variable or measure of effect. In this case,

extreme groups are compared for the same reason, but another consideration also pertains for making a comparison between extreme groups. It is this in many sociometric and psychometric scales we find them highly discriminating at extreme values but less reliable over their middle ranges Thus, there are two reasons for starting ex post facto studies with extreme groups such as quartiles or deciles on the measure of effect

Religious Activity and Personal Adjustment Among the Aged

This is a study of present personal adjustment (spring of 1951) following upon prior experience in religious activities and other factors of a secular nature, among 219 aged persons, 65 years of age and over, who resided in five homes for the aged and two public institutions not exclusively for the aged, in the Minneapolis and St. Paul area.[15] In this study no universe was defined, other than the population of aged persons in these seven institutions, and no random selection of cases could be made for the sample, because of the usual difficulties[16] of interviewing aged persons Chi-square tests were made, however, of differences between the population not interviewed and the sample interviewed, for two of the homes for which data were available in the house record on ten personal and social factors, and on eight of these factors the probabilities were more frequent than the 5 per cent level.[17]

The null hypothesis of this study was· the present personal adjustment scores of aged persons is the same for two groups, one previously participants in religious activities, and the other nonparticipants This general hypothesis was broken down into five specific hypotheses using as the chief independent variable five different measures of previous religious participation such as· present church members versus present non-church members; church members versus those never church members, high religious activity score (for the past) versus low religious activity score (for the past);

[15] Moberg, David O, *Religion and Personal Adjustment in Old Age*, Ph D Thesis in Sociology, University of Minnesota, December, 1951, pp 167

[16] *Ibid*, p 24

[17] *Ibid*, p 32

former church leaders versus non-leaders; and religious believers versus non-believers.[18]

Of these subsidiary hypotheses, the only one which failed to show statistically significant differences on present personal adjustment, was the first hypothesis for which the contrast was made between members and non-members There appeared to be no relationship between present measured personal adjustment and church membership past or present. Owing to limitations of time and space, we shall not attempt to describe the results in detail of all the remaining four null hypotheses, which were not confirmed The one chosen for elaboration herein is the present personal adjustment of aged persons previously participants in church activities is the same as the adjustment of persons who were non-participants.

To test this hypothesis two sociometric scales were needed · one to describe the personal adjustment of the aged as a measure of present effect of factors acting in the past; the other, to measure the degree of religious activity in the past Personal adjustment was measured by the Cavan-Burgess-Havinghurst-Goldhamer scale developed at the University of Chicago and published in 1949. This scale deals with eight areas of living health, friends, work, economic security, reliance upon religion, usefulness, happiness, and satisfactions in family life. The seven items in each area are questions designed to invoke responses which indicate the subject's feelings of satisfaction with respect to the area A total score on all eight areas may be obtained, and also scores on each area separately. It is essentially a psychometric instrument but has been tested for reliability and validity and compares favorably in these respects with other psychometric devices[19]

The second scale required for this part of the study, namely, a

[18] *Ibid*, p 132

[19] *Ibid*, pp 158–159, and Ruth S Cavan, E W Burgess, Robert J Havinghurst, and H Goldhamer, *Personal Adjustment in Old Age*, Science Research Associates, 1949 For example, Sect D, Economic Security, has these items 1 I am just able to make ends meet, 2 I have enough money to get along, 3 I hav'nt a cent in the world, 4 All my needs are cared for, 5 I am provided with many home comforts, 6 I have everything that money can buy; 7 I have to watch how I spend every penny Each question was answered: Agree, or disagree or uncertain, with a weighting system provided to score the responses

sociometric instrument to measure past religious activity, was developed by Dr. Moberg himself It consists of ten items of a behavioral nature, that is, activities performed in the past in connection with a religious institution. The items thus report behavioral facts, not how the subject felt about religious activities. These overt activities include information on membership, attendance at religious affairs, positions and offices held in the church, and frequency of performance of certain religious acts such as listening over the radio to church services, reading the Bible, private prayer, family prayer, saying grace at the table, etc These details are mentioned to indicate more clearly that matters of fact rather than feeling states or satisfactions were used to give this scale some degree of objectivity, as a measure of the chief independent variable of past behavior required to test the hypothesis. Table 44 shows the ranges of scores on these two scales, and indicates clearly a substantial degree of correlation between them

Earlier in this section a limitation of psychometric scales was noted, i.e, the fact that although such scales discriminated well at extreme values, they were often non-discriminating over their mid-range calibrations [20] This is not the time or the place to discuss the reasons and the remedies for this limitation, but it is mentioned now as a reason for the procedure adopted at this point in the present study. The assumption was that if two groups of aged were compared, one showing high scores (over 12 points) on the religious activity scale, the other showing low scores on this scale (under 8 points) there would not be found statistically significant differences on the personal adjustment scale between the two groups, provided there was any truth in the original hypothesis

Proceeding on this basis, it was found that the mean personal adjustment score was 31.1 for 86 persons whose religious activity scores were above 15, and was 18 3 for 41 persons whose religious activity scores were below 8. These results do not support the null

[20] In the present case it turned out that the rating scale on previous religious activities was quite discriminating at its mid-range (an unexpected result), since the scatter diagram of this range shows clearly a substantial degree of linear correlation See chapters vi and xi

TABLE 44 Mean Personal Adjustment Scores by Religious Activities Scores [a]

Religious Activities Scores	Total	Number of Persons Church Members	Non-Church Members	Mean Personal Adj Scores [b] Without Sec E (Religion)	All Sections
23	0	0	0	——	——
22	0	0	0	——	——
21	7	7	0	27 571	33 571
20	4	4	0	28 250	34 250
19	6	5	1	30 333	36 167
18	15	11	4	22 333	28 067
17	14	10	4	25 786	31 643
16	18	16	2	24 333	29 778
15	22	17	5	25 318	31 000
14	19	15	4	22 158	27 579
13	18	12	6	20 000	25 222
12	16	9	7	20 938	26 375
11	17	11	6	18 882	24 176
10	10	6	4	18 800	23 300
9	12	4	8	20 583	25 333
8	8	2	6	12 625	17 250
7	10	0	10	17 000	21 100
6	9	3	6	17 000	20 667
5	1	0	1	15 000	16 000
4	6	0	6	12 667	15 667
3	5	0	5	13 400	15 200
2	1	0	1	15 000	17 000
1	1	0	1	14 000	14 000
0	0	0	0	——	——
Total	219	132	87	21 288	26 352

[a] Taken from Table IX, p 83, of David O Moberg, *Religion and Personal Adjustment in Old Age*, a Ph D Thesis, University of Minnesota, December, 1951

[b] High scores indicate good adjustment

hypothesis Since in this comparison, however, the scores on personal adjustment include area E of the scale, the area that deals with satisfactions in religion, this area was then eliminated in the determination of the general personal adjustment scores, as a result the high group showed a corrected score of 25.3 compared with a

score of 14.9 for the low activity group Evidently the association between personal adjustment and past religious activity was not spurious (due to a common religious factor)

The next problem involves seeking an answer to the question are there secular factors which produce high or low personal adjustment, so that if we control these factors, the former association between personal adjustment at present and past religious activity would vanish? This is a question of partial correlation which we cannot handle except by matching, since many of the secular factors are described only by dichotomies, trichotomies, etc. and not as continuous variables. Data on the following secular factors were available. self-rating on health, present employment, marital status, organization memberships, education, and sex. All of these factors (except present employment), were found to be positively associated with scores on present personal adjustment For example, in the case of organization membership [21]

(1) not a member now, none at age 55 = 25.0 pers adj. score
(2) not a member now, some at age 55 = 26 7 " " "
(3) not a member now, but at age 55
and held an office, = 28 7 " " "

It was therefore decided to control these secular factors by matching [22] As far as possible cases were paired, but when two or more cases occurred in the high score group to one in the other for a sub-class, one case was retained from the high group which matched on some additional factors with the single case in the low group, and the other high score cases were discarded Matching in this manner left 19 individuals in each group, for whom the mean personal adjustment score of the high group was 23 0, and of the low group was 13 0, with the difference statistically significant at better than the one per cent level Evidently variability in the seven secular factors did not explain the association between present personal adjustment and past religious activity. The alternative hypothesis is

[21] Moberg, *op cit*, pp 81–83, also, "The Christian Religion and Personal Adjustment in Old Age," *A S R*, February, 1953, pp 87–90

[22] *Ibid*, pp 50, 155

again supported by the results. These findings do not, however, clarify the relationship to the extent of indicating which came first, potentiality for good personal adjustment becoming a fact in old age, or previous religious activity transforming poorly adjusted younger persons into well adjusted aged presons The specific causal relationship has not been established by this study.

The Logic of Ex Post Facto Design

In concluding this section, it may be worth while to attempt a formulation of the logic which guided the procedures. Each study began with the assumption that our knowledge of causal factors in human relationships could be clarified if a present factor of contrast was regarded as an effect of antecedent factors. In each example the contrast on effect was described by a rating device either in the form of objective facts or by contrasting scores on some objective scale of measurement Then the incidence of specified antecedent factors which characterized each group was explored to ascertain if these factors showed a differential occurrence corresponding to the present differences on the measure of effect.

In all examples, some significant differences were obtained on some of the associated antecedent factors But the validity of using probability tests to determine the statistical significance of the differences, although noted, was not discussed It becomes necessary, therefore, to comment upon the justification for invoking probability as a measure of the significance of the obtained differences. As stated earlier, our process of reasoning in this respect was purely analogical. Science is always skeptical of analogical reasoning, even when, as in these examples, no attempt was made to draw inferences to any larger population Certainly the results as regards the limited groups compared speak for themselves, subject always to the additional restriction that there may be unknown factors at work For example, in the old age personal-adjustment-prior-religious-activity study, there may exist, as unknowns, some common antecedent factor or factors, which explain both the differences in personal adjustment and the prior religious activity differences, each being the

effect of the unknown factor acting in the past. The data of the present studies seem inadequate to solve this basic problem If replications yielded similar results, there would be additional empirical evidence to support the alternative hypothesis

In default of evidence from replications, we have some slight factual evidence from other studies which bears upon the logical problem of using probability in such studies In designs which compare non-random samples and obtain statistically significant differences, the results may be due to the fact of non-randomness alone, or may be indicative that the samples were drawn from different universes. If we turn the procedure around and select two non-random samples from populations *known* to be different, do we obtain statistically significant differences on other factors believed to be correlated with the dichotomous selective factor, or on factors believed to be independent of the dichotomous selective factor? To test this question we selected a sample of 50 census enumeration districts from a "good residential area"[23] and found the mean rental as of 1934 to be $38 10, then we selected 50 enumeration districts from "blighted" or "slum" residential areas and found the mean rental to be $18.13. The obtained difference in rental was very large, $19.97, and the t-value of the difference was 6 38 Next, from the "blighted" or "slum" areas we obtained for a group of 34 low class white families a mean social status score of 39 7, and for a group of 121 lower middle class Negro families, a mean social status score of 60 1[24] This differences showed a t-value of 3 1 The random sampling distribution of rentals of residential properties is a symmetrical distribution approaching the normal curve[25] A sampling distribution of mean social status scores also approaches the normal curve Thus, in both examples non-random samples chosen from populations

[23] In Minneapolis, Minnesota

[24] The measure of social status used in this study was "The Social Status Scale" which rates articles and objects observed to be present in the living room of a home The original standardization of the scale is described in my book, *Contemporary American Institutions*, Harper & Brothers, 1935, and Wm C Brown, 1946, pp 373–397, and pp 61–63 and Appendix A of this book

[25] Schmid, Calvin F, *The Social Saga of Two Cities, op cit*, p 210

known to be different showed statistically significant differences on a related factor. This was to be expected if a strong degree of association existed between the dichotomous factor and the "independent" variable. To test the statistical significance of the obtained differences we entered the normal probability table using estimates of the standard errors based on the non-random samples. This procedure was followed because of the *analogous form* of the respective sampling distributions and the normal probability curve In the ex post facto studies previously reviewed, the logic of expectation and the logic of probability differ from the foregoing illustrations chiefly in that the independent variables are factors or influences operating in the past at different dates.

It would seem, therefore, that although the general logic of a regression equation pertains as a possible mathematical model, the situation in ex post facto studies is more complex, since the independent variables have each a time lag, and while some of them may be additive, others are possibly multiplicative, so that a highly complex multiple regression equation is indicated. To put the matter more systematically, the empirical evidence of these ex post facto studies indicates: (1) quantitative variables are difficult to obtain, whereas qualitative variables or dichotomies or trichotomies, etc, occur frequently For example, sex, marital status, presence or absence of children, employment, disease, etc, are related to personal adjustment as the dependent variable; (2) a time lag factor assumes considerable importance; employment precedes marriage, marriage precedes divorce or separation, and precedes children, etc.; the age at onset of disease may differ, and (3) these "independent" variables are intercorrelated; the presence or absence of children reacts upon marital status, and is related to divorce or separation; age at onset of disease may be a function of size of family, etc Starting with these facts and turning now to the problem of an appropriate model, it appears at present that little can be said, except (1) that the ordinary or "classical" regression equation is not appropriate; a more powerful and versatile formula seems needed; (2) stochastic difference equations suggest one type of model, since they embrace the time lag factor and offer sets of systems of equations;

$$\sum_{j=1}^{r} \sum_{k=0}^{p_{ij}} a_{ijk} X_{j,t-k} + a_i = \epsilon_{it},$$

but (3) stochastic difference equations unimpeded by linearity restrictions seem needed particularly

$$X_{1,t} = a_i X_{2,t-1} (X_{2,t-2} + b_i) + c_j X_{3,t-1} + \ldots + r X_{n,t-n} + Z_{ij}$$

Andrew W Marshall's paper, "A Large-Sample Test of the Hypothesis that One of Two Random Variables Is Stochastically Larger than the Other,"[26] wherein it is shown that two groups of patients have different ages at onset of disease, is also suggestive of a method of statistical analysis appropriate to the costs of medical care of the aged we have before us.

Of the ex post facto studies summarized herewith, only the study of the cost of medical care of the aged would seem appropriate for analysis of variance. In this study the initial random sample of 354 cases could have been broken up at random into 35 small samples of 10 cases each and then analyzed But since we lack information about the larger populations of individuals who existed 5, 10, 15, 20, etc. years ago, of whom our present 354 persons were a part at these earlier dates, we are unable to test the randomness of the 354 cases as of those earlier dates. Empirical evidence drawn from life table trends points to the present 354 cases as a composite of survivors *after* selective influences on past populations, whose original total at birth is estimated at approximately 2476 persons of both sexes Even this line of reasoning takes no account of mobility, since there might be living in 1950, elsewhere than in Hennepin County, some members of these populations Table 45 shows in skeleton outline the estimated original populations by age groups and sex of the present 354 survivors in the random sample of 1950. Our estimate of

[26] *J A S A*, September, 1951, pp 366–374, H B Mann and A Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, July–October, 1943, pp 173–220, and J V Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, 1937, pp 161–181. "Symposium on Stochastic Processes," papers by M S Bartlett and D G Kendall, *Jour. Royal Statistical Society*, 1949, pp 221–229, 230–265, also J A Jahn, "Field Experiments in Sociology," *Research Studies*, State Col Wash, 15 162–168, and A T Read, "Stochastic Processes and the Study of Growth Phenomena," *J A.S A.*, June, 1954, p 331.

TABLE 45 Survivors by Age Groups and Sex (White)[a]

	65–69	70–74	75–79	80–84	85–89	90–94	95–100	N
At 1950	M F	M F	M F	M F	M F	M F	F	
Survivors	17 38	33 69	36 47	27 45	14 13	3 9	3	354
1941–	20–43				34– 79			
1945								
(60–65)		51– 97				27–102		
			85– 92				103	
1921–	27–52			118–150				
1925		70–119			149– 97			
(40–44)								
			115–113			118–342		
1901–	29–55			160–185			347	
1905		74–131			202–119			
(20–25)			123–120			159–420		
1881–	*32–59*			171–189			426	
1885		*82–134*			217–125			
(at birth)			*135–128*			171–442		
				187–208			449	
					237–134			
						187–473		
							480	
Totals at Birth[a]	32–59 91	82–134 216	135–128 263	187–208 395	237–134 371	187–473 660	480	2476

Diagonals show survivors of estimated original populations at ages 60-, 40-, 20-, etc Dates at left apply only to frequency entries in next column at right, remaining frequency entries staggered in diagonals to indicate diminishing sizes of populations

[a] This estimate of 2476 persons in the original generations is based upon calculations in round numbers for the ages 65, 70, 75, etc using the life table for 1939–1941 of the United States Department of Commerce, Bureau of the Census, *Vital Statistics—Special Reports* (vol 19, no 4) and allowing for sex differentials, and applied to the age and sex breakdown of the sample of 354 persons (as per Stinson, *op cit*, p 21)

the original total populations as 2476 persons is based upon calculations in round numbers for age groups, using life tables of the United States for the years 1939–1941 from the United States Department of Commerce, Bureau of the Census, *Vital Statistics—Special Reports* (vol. 19, no. 4) in which each sub-group of survivors (by age and sex) was a group of those then living as a proportion of an original population of 100,000 at birth. Although the figures are rough estimates, they do give some idea of the sizes of

samples that would be required at intermediate years, and also some idea of the numbers lost by mortality (but not by mobility)

In short, there seems little that we can do with this random sample of the presently living aged, to ascertain how these represent the total populations at different past dates, of which populations these are the survivors It is considerations of this kind which state the sampling problem to be solved before one can justify the application of modern statistical methods that rest upon assumptions which our data do not seem to meet Since this is the situation for the medical cost study, it is even a greater limitation upon the other examples, wherein even the present group could not be randomly selected.

Chapter XI

SOME PROBLEMS IN PSYCHO-SOCIAL MEASUREMENT

SOME of the complications and difficulties which hamper the social scientist in his search for objective methods of description and measurement of human behavior in the free community situation are. (1) popular pressures to "do something at once" about extreme individuals who are thought to be responsible for acute social problems, individuals who are criminals, juvenile delinquents,[1] psychopathic persons, etc, *before* the social scientist has accumulated a sufficient volume of verified knowledge to justify using his specific findings as a basis of social action and social policy; (2) social and psychological obstacles[2] to acceptance of the existing fund of social science knowledge; (3) the fleeting and intangible character of such important phenomena as social attitudes and opinions, and (4) the tendency of human subjects of research to falsify their responses to honest inquiry or to resort to substitute response mechanisms of compensation, projection, rationalization, and so forth. This chapter will offer a brief description of these problems

EXTREME INDIVIDUALS

Individuals whose behavior is an extreme departure from the expected pattern of behavior have always been of great interest Sometimes the behavior of such extreme individuals has had a catalytic influence upon members of a social group, or such individuals, acting as leaders, may stimulate others to effective social achievement.

[1] See chapter IV, pp 84–94, for a description of a program for the prevention of juvenile delinquency based on the findings of systematic research

[2] The description of these obstacles was published originally as an article by the author, "Social Obstacles to the Acceptance of Social Science Knowledge," *Social Forces,* October, 1947, pp 7–12, see also George A Lundberg, "Alleged Obstacles to Social Science," *Scientific Monthly,* May, 1950, pp 299–305, and "Semantics in International Relations," *A Review of General Semantics,* Summer, 1949, pp 333–337.

Many innovations in culture that are highly valued came from men of genius[3] On the other hand, extreme individuals may be a burden on society, to be supported or segregated, as chronic relief cases[4] or the feebleminded.[5] Epileptics,[6] schizophrenics,[7] and drug addicts[8] are often associated with situations of conflict and tension in families and other social groups But all such individuals are relatively rare occurrences. To devise special treatment for them or to protect society from them, research on random samples of the general population with assumptions of normal distributions of some measure is not appropriate, except to select control groups for comparisons This prevailing concern with deviates from those behavior forms generally regarded as normal or desirable is supported by moral and ethical ideologies, and embodied in tradition Deviates such as delinquents, defectives, and psychotics are regarded as social problem-makers To check such deviates in the interest of maintaining social stability, efforts are put forth by the ameliorative and curative professions of medicine, nursing, and social work To try to restore such deviates to behavior considered normal or more desirable is an expression of democratic ideology Hence arise pressures to "do something about it"

Scales of measurement are needed to yield quantities[9] that describe deviates or behavioral extremes, and even specialized sta-

[3] Frederick Osborn, *Preface to Eugenics,* revised ed , Harpers, 1951, p 72 Osborn reports estimates on persons of genius of 0 6% for I Q over 140, and a small fraction of 1% for I Q over 180

[4] Bradley Buell and associates, *Community Planning for Human Services,* Columbia University Press, 1952, p 9 Buell reports that 6% of the families of St Paul, Minnesota, were suffering from multiple problems and were absorbing over half the combined services of the community This 6% figure has since been confirmed for several other communities studied

[5] Osborn, *op cit ,* p 31 Two per cent of the population is estimated as feebleminded

[6] *Ibid* Epileptics are estimated at 2 9 per 1,000

[7] Osborn cites Kallmann's estimate that schizophrenics are 0 85% of the general population

[8] Drug addicts have been estimated at 3 per 100,000

[9] As, for example, the Guttman scaling method to be noted presently, and the Minnesota Multiphasic Personality Inventory (MMPI) for the measurement of departures from normal or expected behaviors

tistical methods of analysis may be helpful, such as extreme-value theory[10]

Extreme-value theory in statistics, as developed by Emil J Gumbel[11] and Julius Lieblein,[12] is concerned with the largest or the smallest values in a set of magnitudes, and does not concern itself with average values and other quantities which follow a symmetrical or normal distribution For the largest observations, for example, the distribution is skewed to the right, and very large values have a much greater chance of occurring than very small values. The method has been applied in the study of droughts; the fracturing of metals, textiles, and other materials under applied force; and in fatigue phenomena. Since the social scientist is often called upon to study extreme forms of human behavior that are thought to jeopardize the cohesion or solidarity of social groups, we suggest that the application of extreme-value theory to problems of human relations in social groups may be fruitful, although limitations of space and time, unfortunately, have not permitted us to attempt any adaptation herein.

Social Obstacles to the Acceptance of Existing Social Science Knowledge[13]

Application of existing social science knowledge to the amelioration of problems of human relations is hampered by existing habits of thought and action Curiously enough, this resistance is directly proportional to the soundness of this social science knowledge; the principles of human relations inductively derived from empirical research and observation of human behavior are less likely to be

[10] E J Gumbel, "Some Applications of Extreme-Value Methods," *The American Statistician*, December, 1954, pp 14–17 (a publication of the American Statistical Association) See especially for extended treatment E J Gumbel, *Statistical Theory of Extreme Values and Some Practical Applications*, National Bureau of Standards, Applied Mathematics Series 33, Feb 12, 1954

[11] *Statistical Theory of Extreme-Values and Some Practical Applications*, National Bu of Standards Mathematics Series 33, 1954

[12] *A New Method of Analyzing Extreme-Value Data*, NACA Technical Note 3053, Nat'l Advisory Committee for Aeronautics, January, 1954

[13] F Stuart Chapin, "Social Obstacles to the Acceptance of Existing Social Science Knowledge," *Social Forces*, October, 1947, pp 7–12

accepted by intelligent persons than are prevalent panaceas and superstitions

At least eight social obstacles[14] to the acceptance of existing social science knowledge of a sound nature may be noted· the subject matter of social science (crime, sex, politics, wealth, etc) arouses emotional reactions and diverts attention *from* an analysis of the qualities of the objects observed *to* an expression of how we feel about these objects; a normative set or a value-judgment rather than an impersonal approach to problems of social relations is traditional; the observer is himself a part of the social situation or the social relationships he tries to observe, much significant sociological information has a confidential and privileged character; widespread conspiracies of silence exist and prevent effective decision-making about acute problems; unpleasant consequences to minority groups usually follow the application of sound principles of human relationship, much of the subject matter of social science is purely verbal or language behavior and hence seems intangible; and an understanding of the problems of human relations requires the concentrated, consistent, and sustained mental concentration of a new language which few are willing to learn. Let us now consider in more detail each of these obstacles.

1 The Subject Matter of Social Science Is Emotion-Arousing

Thoughtful leaders of modern communities express concern about such problems of human relations as criminal behavior, juvenile delinquency, divorce, prostitution, maladjustment of sex relations, injustice to weaker members of the community such as orphan children and working women and minority race groups, strikes and lockouts, monopoly, cut-throat competition, political corruption and spoils. Now it is a common property of all these problems of human relations that the normal and first response to them is an emotional response rather than an intellectual and an impersonal reaction Confronted with any one of these problems it is difficult to make impersonal judgments and to take a detached point of view.

[14] Individual obstacles in the pathological mechanisms of abnormal psychology are not specifically treated herein, nor autonomic response systems

Conventional morality and orthodox religion insist upon an attitudinal stance with respect to all these problems and condemn an objective approach. Thus the social climate in which we see these problems tends to reinforce a natural tendency to respond to them with anger, fear, scorn, greed, loathing, pity, affection, sympathy, loyalty, reverence, or awe, as the case may be A curiosity response which seeks to observe, describe and record, rather than to express how the observer feels about the problem, is an unconventional response and excites suspicion and disapproval A. G. Keller, in stating the contrast between the attitude of the natural science observer towards his subject matter and the social scientist towards his subject matter says, "A man can count the legs of a fly and not have his heart wrung by finding them too many or too few!"

Thus it is that emotion-arousing subject matter diverts attention from the characteristics of the subject matter to a preoccupation with how one feels about this subject matter And the impartial social observer comes to be regarded as hard-hearted, callous, and cynical. As a matter of fact, a cynic is usually merely a bad name that a neurotic calls a realist whom he does not like; and a neurotic is a person whose habits of response tend toward substitute response systems, rather than toward direct response to direct stimuli; and such response systems have self as the center of reference. In the neurotic person, responsibility for his acts and decisions is avoided by escape into lengthy explanations in which words are used which have no verifiable fact-referents.

2. *The Normative Set and Value-Judgment Approach*

Emotion-arousing subject matter tends to develop habits of response which either consciously or unconsciously attach "praise" and "blame" to purely natural situations in the community of human relationships "Who is to blame?" we inquire, and not, "What is the cause?" Use of the pronoun "who" indicates a personal and anthropomorphic explanation based on self-reference. Use of the verb "blame" clinches the matter as one of bad motivation, and thus adroitly shifts responsibility to another. Whereas to ask, "What is the cause?" is to use a less emotional expression, in which the interrogative pronoun "What" refers to a thing and is substituted for a

personal pronoun, and calls for selection from an indefinite number of factors. Finally, the noun "cause" is entirely devoid of personal implications, hence is neutral, seems colorless, and is not interesting to most persons

The value-judgment approach is natural for the masses of people and is the normative approach of the "reformer" But for purposes of systematic and sound knowledge of human relations these uses of language are misleading, since they merely reveal the feelings of the speaker in reacting to the problem and do not describe the problem in transmissible and recordable terms Although the statements made may be expressed in the grammatical form of assertions of fact they deceive both the speaker and the hearer as to what is really taking place, since value-judgments are expressions of desire, of what is wished for or disapproved, and contain terms which do not describe the qualities of the thing or situation under observation

But unfortunately normative statements have their own psychological utilities. They provide a simple principle of classification (all things are put into dichotomies) as a substitute for the bewildering variations of real experience. Thus anything is "good" or "bad", we are favorable to it, or we disapprove it. In an increasingly complex social world the normative categories simplify the adjustment process by reducing the number of verbal signs required in thought or in conversation. They also save the fatigue of critical analysis and close observation to details They seem to provide rules which can be relied upon to promote a sense of security in a threatening world. For the slightly more sophisticated, they seem to protect one from the seeming uncertainties of the calculus of probability, with its realistic admission that there can be no absolute precision in predicting events To try to think through all of these principles is very confusing to the average layman

3 The Scientific Social Observer Is Himself a Part of the Social Process He Tries to Observe

The student of human relations has few tools of objective observation which are mechanical instruments of precise measurement. He has to resort to words to describe language behavior, and to

these words his whole organism is often strongly conditioned from early childhood Participation in the give and take of the surrounding system of communication rests upon habit patterns that are well established, and with these go an attitude structure which tends to set the approach to screen out those social stimuli which are displeasing. Thus are unpleasant consequences ruled out and painful adjustments avoided And since each of us has as many roles as the numerous groups with different ends to which we belong, internal tensions and conflicts are resolved by subdivision of personality into separate compartments, and this prevents a wholesome integration of ourselves.

Since the days of the Greek philosophers men have sought to escape the suffocating and all-surrounding social climate of conventional morality and tradition by finding refuge of a sort and a certain degree of detachment about human relations which comes from the study of archaeology, history, and ethnology. In the study of the cultures of other times and of different places some appreciation is gained of the recurrence of problems of human relations and of the relativity of the norms used to judge the importance of these problems in any given time and place. Unfortunately this escape into detachment is usually a mere translation from one folk language into another, either as between different periods or different regions It all remains on the same plane of meaning—nominal meaning No real emancipation from the folk pattern is achieved. No penetration by aid of semiotic analysis or by social-psychological study is effected One is still bound by nominal relationships—by the names given to relationships It is still assumed implicitly that, since certain words customarily stand for real things, there must be somewhere an existing thing for which every noun stands. The question whether there are terms which merely represent other linguistic forms and never can be shown to stand for an existing thing, is a question seldom raised and seldom faced Only in recent times has the idea of an operational definition been grasped the principle that a concept of human relations may be defined in terms of the operations performed in measuring it. Take the concept of "morale," certainly a social concept of some importance in the human relations

of both war and peace, it has been only recently that psychometrics and sociometrics have given us a measure of this factor in human relations; and these measures used to implement social policy

4. The Confidential and Privileged Character of Much Sociological Information

Much sociological information is based upon observations and inferences about particular personal relationships in specific social situations. There are recognizable mechanisms and patterns of individual behavior in social situations which involve ascendance or submission, shyness or boasting, escape into verbal fantasy or courage in facing unpleasant facts, which is learned by long experience in participant observing of social organization Those who have had extensive practice in committee work, conference procedure and group pressure tactics, develop an insight into human motives and behavior, which is difficult to reduce to measured and quantitative terms but which may none the less supply a very effective understanding of what is going on. This shrewdness has much more depth than a superficial "being in the know" Much of this experience involves recognition of "the psychological moment" promptly as it appears, and there is also quite probably an element of good luck in being "on the spot at the right moment." The most effective use of such empirical social knowledge is to exercise it as a "wise counselor" rather than as an "expert," which often is offensive to others, since every man believes that he alone is a good judge of character and can read the signs of the social times with insight and sureness

Thus the sociologist who is an active participant in the life of the community and not "an armchair theorist" accumulates a considerable volume of information about specific leaders, their weaknesses and their strengths, which has to be kept confidential since it can be very "dangerous knowledge" to him. In fact, the type of participant observer described can, if he is indiscreet, i e , voluble and talkative, actually "know too much" for his own good. Now this confidential and privileged character of some sociological information makes scientific (objective, public) formulations of knowledge

difficult, despite the fact that such information may be of real strategic importance. In sound social administration and executive leadership, however, "secret" knowledge is essential to the smooth operation of the social organization The good executive has to "keep his own counsel" and confide in no one The essential point is to keep it secret that you have a secret; otherwise such a leader is open to attack, raids, and reprisals Now this kind of social knowledge is far from being scientific, since it is not transmitted or recorded, and yet it may be very effective Herein we have another obstacle to the development of a genuine science of human relationships, since science is essentially public, transmissible, and recordable. When an effort is made to objectify such information by the device of anonymity, it tends to lose its authority Perhaps this dilemma never can be resolved.

5. *Conspiracies of Silence*

Conspiracies of silence complete the picture and complicate the process of making a decision about acceptance or rejection of the known principles of social science as a rationale for individual and group action, and often take the edge off rational effort

There are "conventions of good taste" which prevent asking questions that may be embarrassing to another person because such questions are interpreted as prying into his personal affairs. In sociological questionnaires the inquiry is worded to avoid being unduly inquisitorial. Solenberger found in a study of homeless men that the question, "Are you married?" brought answers that seemed to be in smaller proportion than seemed reasonable When he rephrased the question, "Where is your wife now living?" he obtained a larger proportion of married men in the group investigated Good taste prevents interference in the affairs of a neighbor or a relative when the evidence clearly points to the fact that a promising child is being badly spoiled Conventions for the protection of personal privacy take the forms of rights; freedom of thought, freedom of speech, freedom of religious worship. These have been won at great cost. But their anti-social consequences have seldom been frankly faced

Often fear of reprisals from powerful and intrenched interests curbs the impulse to expose the existence of undercover political corruption, criminal rackets or even of unethical practices, not so much with the idea of personal safety, as from the knowledge that such an exposure would jeopardize further inquiry and fact gathering People are hesitant about "stirring up a mare's nest"

Conventions of line organization require reporting only to those next above, not going over the head of immediate superiors, and being loyal to the organization and its leaders, this can be carried so far as to block new ideas and useful criticism, and to thwart self-expression. Such conventions often stand in the way of sociological research and application of existing knowledge about social inventions.

6. *Unpleasant Consequences to Minority Groups When Social Science Knowledge Is Applied*

Acceptance of the methods of social science and application or existing sound knowledge entails disagreeable results The reason is that these methods and principles supply a rationale for thought and activity which is unpopular with special interest groups.

Economics provides logical statements and reasons why progressive income and inheritance taxes are a dependable basis of public revenue, and this provokes the wealthy classes, regulation of competition and control of monopoly promote the elasticity and productive capacity of the economic system, and this exasperates certain predatory business interests; responsible leadership in collective bargaining arouses the enmity of inexperienced or racketeering labor leaders, free trade seems to clip the wings of predacious commercial interests who wish to use high tariffs only to protect uneconomic infant industries, wages are not only income but are costs in some other price structure, but any regulation of wages in an inflationary period arouses the resentment of organized labor; etc., etc

Political scientists have shown the reasons why a merit system in the selection of public servants may provide competence in performance and elasticity in government functions, but this encounters

the opposition of political spoilsmen on the one hand and on the other the enmity of entrenched bureaucrats; periodic checks on the quality of food commodities, the safety of buildings and plumbing, the adequacy of fire protection devices, by an adequate and competent staff of inspectors meets the opposition of taxpayers leagues and the hostility of "fixers"; new laws do not eradicate social evils permanently and in short order, but depend on a wide base of public acceptance and conformity to achieve results, which invites the scorn of starry-eyed social reformers and the grumbling of an expectant public; etc, etc

Sociologists provide logical reasons why: new laws and social inventions to improve social organization do not so much supplant antiquated social machinery as they add to the total of existing regulations, which excites the disbelief and animosity of eager reformers and irritates the public who were given to expect quick relief and a simplification of the pressures of regulatory measures; class prejudice and race prejudice rest upon the insecure foundation of a myth of desired superiority rather than upon the realities of proven inferiority, and this undermines cherished beliefs, shakes the very foundations of the social status system, and is resented as "dangerous" teaching; poverty and crime are often not so much what we think they are, as that they are only what we think, since in the case of poverty, the higher the standard of living the more ways there are of being poor; and in the case of crime, what for the most part is regarded as crime is a matter of the conventions of a given time and place. Such statements seem to strike at the rock of ages and tend to enrage the orthodox, who see in them a threat to the security of their fondest fancies. Etc, etc.

7. *Much of the Subject Matter of Social Science Consists of Verbal Behavior—Intangible and Trite*

Attention to the immediate subject matter of social science reveals the fact that much of it consists of language behavior, oral or written. There is the oratory of the demagogue; debate which resorts to the dialectic of concepts to win a point rather than to clarify an issue, controversies over panaceas and pure fantasies,

ideological doctrines of a social, political, or religious nature; fruitless discussions in committees and in conferences, voting behavior in which many mark a ballot without adequate knowledge of the relative fitness of different candidates for office; editorials in newspapers, journals of opinion, tracts and pamphlets, volumes of advertising matter, etc , etc.

Compared with this the subject matter of physical science, atoms, organisms, and stars, exists independently of language and in its own right, with a hard material integrity It is no wonder then that the corporeal achievements of chemistry and physics tend to blind the masses to the real fact that scientific method rather than the machinery of the laboratory is the greatest value of science

8. Concentrated and Continuous Mental Effort Is Required to Know that such Concepts as Intangibility, Relativity, and Probability Are Basic to Social Understanding

Continuous mental effort and close attention to detail are necessary to scientific understanding of the problems of human relationship; this is the case because the regularities are hidden in intangibility and relativity, and the predictions are in terms of probability.

Unless embodied in a new language of social concepts, the regularities of human relationship are lost within the intangibility of phenomena whose points of reference in space and time are in continual movement so that they elude our grasp and fade away. What is "morale," "public opinion," and "inflation"? To observe, describe, and record these phenomena of human relationship, the folk language of common terms is wholly inadequate But to gain acceptance of the dependability and utility of operationally defining these social phenomena is to educate the layman in the validity of such systematic methods of study as are afforded by the use of psychometric and sociometric scales, public opinion polls, and the careful application of statistical research Here we get involved in a special vocabulary of new and strange terms which excite distrust, despite the fact that the public is willing to accept on faith such other terms as isotopes, neutrons, genes, carbon rings, and so forth

Now the use of psychometric and sociometric scales have given us

some population norms We can tell, for instance, how much below or above the population norm of morale a given unemployed person is. We can describe the different opinions on a social issue like price control held by rural and by urban people And the results of a given election have been predicted within an error of 0.7 of 1 per cent from a polled stratified sample weeks in advance of the event. Nevertheless these norms are always relative to a given population in time and place. As yet no universals have been discovered in empirical social science We are still dismayed by the activities of Russians and Japanese, although we are able to predict the behavior of Anglo-Americans with increasing precision.

This state of affairs irritates the concrete-thinking and practical-minded person no end. He has lived with the habit of mental-effort-saving which relies on dogmatic answers to all questions. Why is not social science knowledge as authoritative as he incorrectly assumes all physical science knowledge to be? Well, we know that real scientists make no such claim, and are in fact most modest, considering their brilliant achievements But the highly technical character and the great volume of scientific knowledge elicits, as it should, a popular respect, because these attributes implement effective procedures in preventing the spread of disease, in controlling insect pests, in producing stronger alloys, and in the facilitation of mechanical means of communication and of transportation.

By contrast, the knowledge of the social sciences deals with the presumably familiar problems of human relationship, for which the man of the street feels a peculiar penchant When, therefore, he is told that real understanding of these assumedly obvious relationships requires thinking in terms of a new and specialized language equipped to deal realistically with a world of really intangible relationships, relationships that always seemed to him self-evident, he is inclined to be skeptical of this new language and consign it to the limbo of speculative thought. To be told that a principle of relativity runs through the structure of social knowledge does not seem to provide a stable basis for the determination of social policy, and disappoints all who desire a "settled basis of thought," "solutions" for the acute problems of human relationship, and "proof"

of the validity of desired goals of effort. The fact that human problems are recurrent throughout recorded history despite some differences in culture, and that amelioration of these problems rather than their solution, is about all that the present stage of social knowledge can honestly promise, does not satisfy the natural desire for a permanent security in human relations

Finally, the acceptance and application of existing social science knowledge demands some understanding of, or at least willingness to acquiese in the validity of the principle of probability. Take the case of public opinion polls. Most persons cannot believe that these polls based upon interviews with almost mythical persons, because so few individuals have ever known someone who was actually interviewed, do nevertheless collect information which yields surprisingly accurate predictions of election events that occur weeks later. The average congressman is also skeptical or more often prejudiced against such polls, so that he responded to the insistent pressures of special interest groups and lobbies (which represent minorities only) and decided to do away quickly with rationing and price controls, when the facts were that frequent public opinion polls showed repeatedly that a large majority (70 per cent and over) of the public were willing to comply with such controls and expected gradual relief from them rather than sudden lifting. The result of this failure to base public policy on sound social knowledge scientifically obtained was the deplorable situation in which the nation found itself about 1946 with respect to rationing policy.

Now what has all this to do with probability? Simply that the modern public opinion poll (the use of which is, by the way, possible only in democratic countries) is based upon interviewing persons who represent all groups of society within the prevailing range of education, income, occupation, and other social stratification factors This representation is made dependable by stratified random sampling; and in random sampling, the principles of probability govern the selection of interviewees Consequently a very small number of properly chosen interviewees will yield dependable results, the reliability of which may be calculated in advance by relatively simple application of the principle of probability, and this

assertion is not merely sound in a theoretical sense; it has been proved true by the most stringent test of verification—prediction of results which do happen

Other applications of the theory of probability are made in sampling studies of other problems of human relationship, and somewhat more technical applications are basic to the whole development of psychometric and sociometric scales used to measure attitudes, opinions, morale, individual adjustment, intelligence, manual dexterity, personality traits and a host of similar instruments of scientific observation which in application remove specific social factors from the limbo of the intangible Beyond this, there is the whole system of study known as experimental designs, which is beginning to show the way to isolating specific cause and effect relationships from the tangled web of human relationships, and here again the theory of probability plays a decisive role

These eight obstacles to the acceptance of existing social science knowledge by no means exhaust the list of barriers to applied sociology, but they may serve to indicate briefly the nature of the complications which are inherent in such phenomena as social relationships, in contrast to the tangibility and externality of purely physical phenomena.

SOCIAL ATTITUDES AND OPINIONS

Let us now consider briefly points 1, 7, and 8 in the aforementioned list of obstacles to the acceptance of social science knowledge, and deal with them as so many facets of a single problem We shall examine them from the point of view of such statistical concepts as a universe or population of data to be sampled by some research design, and then give brief attention to some new developments in theory and method of research which may help to resolve some of the difficulties.

Social attitudes toward other individuals and personal opinions are psycho-social phenomena of considerable practical importance, but they are intangibles to begin with, and fleeting and unstable in addition. Attitudes and opinions change with indoctrination and propaganda, and may differ after new experiences and travel, and

hence are difficult to define in terms of statistical parameters Even if defined for the population of a given date and place such a definition may become highly inaccurate after a brief interval of time

Public opinion polls have often been successful in predicting voting behavior This has been the case when probability samples based on stratified area factors are utilized The universe of people in a given area may be studied by procedures adapted from the experience of probability sampling of physical areas The attitudes or opinions obtained by interviews may yield an accurate picture of the distributions of different attitudes or degrees of intensity of some single attitude, so that reasonably dependable predictions of stability or change have been made for given populated areas

But there are other universes of attitude and opinion concerning which we know much less. For example, how closely or consistently does a person's verbal response, expressed as an opinion in the presence of strangers, reflect his inner convictions about the same issue which he might feel free to express only in the presence of trusted and intimate friends? If we accept the first situation as yielding true results, may we not have missed the real core of the matter? In short, we must postulate at least two universes of opinion one, the public expression of opinion, the other, the private expression of opinion. In other words, the surface patterns of attitude and of opinion may have complex substructures. Here are two universes of a more complex and intangible character than the human population of a physical area They are more complex universes because they are likely to be differently structured From the semantic point of view the verbal response accepted as evidence of opinion may be primarily in the language of emotional expression and only secondarily a response in the language of representation, and vice versa. Cultural differences also appear in the desire to follow the proprieties of the interview situation so as to make a socially acceptable or an expected response, one which is quite unrelated to underlying convictions. But this is a problem explored in textbooks on social psychology.

If these conceptions of two universes of attitude are accepted as a partial analysis of the problem, we then turn to the next issue.

What *kinds of questions* are required to plumb the depths of human attitude structures, and how should these questions *be worded* to elicit spontaneous or true response? Here we find another problem of definition of the universe to be sampled or studied. Again, we do not know a great deal about the universe of questions which would need to be asked so that our research on attitudes and opinions may yield objective data. It should be evident that parametric description of such universes would be difficult to achieve Nevertheless, as we shall see, the new methods developed by Louis Guttman[15] and Paul Lazarsfeld[16] have thrown considerable light upon these problems and pointed the way toward understanding This state has been achieved by the wide use of the Guttman scalograms, in the mathematical proofs supplied by Guttman, and in the mathematical models developed by Guttman, Lazarsfeld, and others.

This chapter is not the place to describe the Guttman scalogram methods, since such description has now found its way into numerous textbooks on social research methods, but it may be useful at this point to indicate briefly the theoretical developments that have followed These supply a much needed and integrated rationale for social psychology as a substantive area of social science on the one hand, and on the other supply a methodological base for social measurements and the procedures of empirical experimental designs in the study of human relations

An elaborate and integrated theory is beginning to develop from the juxtaposition of a considerable mass of data derived from use of the Guttman scalogram methods, and from analytical work on the theory level, which brings together some of the older and some of the more recent developments in factor analysis.

Guttman's *facet theory*[17] deals with sets of elements. It is useful for designing universes of content for research projects. It enables

[15] Chapters by Louis Guttman and associates in *Measurement and Prediction*, vol IV, in *Studies in Social Psychology in World War II*, Princeton University Press, 1950

[16] P F Lazarsfeld, *ibid*, chapters 10 and 11, pp 363–472

[17] L Guttman, "Role Conflict An Example of Facet Analysis," report submitted to the Human Resources Research Institute, United States Air Force, by the Laboratory of Social Relations, Harvard, 1952

great savings in sampling populations and in sampling content items by using parsimonious techniques, and supplies a restatement of R. A. Fisher's approach to design of experiments. It leads to a solution of the problems of "general attitudes" versus "special attitudes" His *radex approach*[18] analyzes the problem of order among variables, much as scale theory studies order among people Sets of scores have an implicit order among themselves, from the least complex to the most complex. This order is called "simplex." But circular orders are found among variables representing differences of complexity rather than of degree, and such variables are called "circumplex" The common factor analysis as developed by Spearman, Thurstone, Holzinger, and Hotelling *et al,* which deals with the structure of correlations among tests, is related to the radex theory Another part of this network of systematic thinking and research in applied mathematics is Guttman's *image theory,*[19] which is a general theory of deviations, with "images" being the predicted values. Here the so-called Israel *Alpha* technique and *Gamma* techniques[20] supply the procedures Then Guttman's theory of *nodes*[21] (ideal scale types) concerns sub-populations of a given universe, in which any two items are statistically independent of each other. It extends common factor analysis to characteristics that the social scientist so often encounters—qualitative data and curvelinear quantitative data. Besides common factor analysis, Lazarsfeld's latent factor analysis is a special case of nodular structure theory, and is concerned with a sub-population wherein each item in an observed universe is statistically independent of any *combination* of other items Finally, Guttman is developing a new *meta* theory which will deal with closed dynamic systems

This far too brief statement of an emerging system of thought

[18] P F Lazarsfeld, *Mathematical Thinking in the Social Sciences,* Free Press, Glencoe, Ill, 1954, chapters 5 and 6

[19] L Guttman, "Image Theory for the Structures of Qualitative Variates," *Psychometrika,* December, 1953, pp 277–296

[20] See chapter 19 in Matilda W Riley, J W Riley, Jr, J Toby, *et al, Sociological Studies in Scale Analysis,* Rutgers Press, 1954

[21] Appeared in a mimeographed form as part of an address before the Annual Conference of the World Association for Public Opinion Research, and of the American Association for Public Opinion Research, April 22–25, 1954

may serve the purpose of indicating to the serious research student the fact that there exists a rapidly developing body of knowledge about interrelated methodological problems of research by scale construction, experimental designs, background analytical systems, and, not of least importance, a new theory of social psychology. It may be noted also that some of the problems we have stated may be resolved when they are viewed against the background of this larger approach, to which many research workers have contributed.

Some Variations in Response Behavior

The student of psycho-social behavior, in his efforts to evaluate changes associated with treatment or differences between groups, is often frustrated by the small size of measured differences [22] This problem was noted in Chapter III, pages 37 and 38, and in Chapter V, pages 104 to 107.

Bias of response to questionnaires and in interviews has long been recognized as a problem,[23] and apart from semantic confusions, cheating, malingering, and intentional falsification are sometimes present These factors may account in part for some of the small differences found in evaluation studies [24]

The problem of bias, malingering, and falsification of response deserves careful analysis and investigation. Since, however, this is a problem more appropriately dealt with in books on scale construction and investigations which use field interviews, we shall merely note here a few precautions taken in certain scales which illustrate the way in which the problem has been handled in specific situations

The Minnesota Multiphasic Personality Inventory (MMPI) is

[22] See, for example, J F Short, Jr's review in *A J S*, May, 1954, pp 387–388, of the study reported by E Powers and H Witmer, *An Experiment in the Prevention of Delinquency The Cambridge-Somerville Youth Study,* Columbia University Press, 1951

[23] G A Lundberg, *Social Research*, Longmans, Green, 1942, pp 46–47, 365, 394, C L Fry, *The Technique of Social Investigation,* Harpers, 1934, pp 64–66, Twila E Neely, *Error in the Interview,* 1937, pp 17–27, Mildred Parten, *Surveys, Polls and Samples, Harpers,* 1950, p 102, and C A Moser, "Interview Bias," *Reviews of the International Statistical Institute,* 1951, pp 1–13, Riley, Riley and Toby, *op cit.*, pp 7, 11, 20, 207

[24] Powers & Witmer, *op cit*

an example This inventory " . is a psychometric instrument designed ultimately to provide, in a single test, scores on all the more important phases of personality." [25] It includes 550 statements covering a wide range of subject matter, and has nine main components. hypochondriasis, depression, hysteria, psychopathic deviate, masculinity—femininity, paranoia, psychasthenia, schizophrenia, and hypomania. From the point of view of our present interest, four subscales may be noted.[26]

The *?* scale number of items unanswered; when score is high no conclusions would be safe

The *L* score. tends to indicate any naive attempt of the subject "to put himself in a good light," chiefly with reference to personal ethics and social behavior.[27]

The *F* score high scores indicate that persons may be attempting (perhaps unconsciously) to show themselves in a bad light, i e. "faking bad."

The *K* score: yields evidence of more subtle indications of defensiveness and lack of candor This scale is signicant for borderline profiles, whose *T*-scores range 65–80

Harrison Gough [28] compared the results obtained from eleven professional persons familiar with the MMPI who took the test in three ways first, as frank self-appraisal, second, in an attempt to simulate a severe psychoneurosis; and third, in an attempt to simulate paranoid schizophrenia Then these simulated records were compared with 57 severe psychoneurotics and 13 paranoid schizophrenics. He found (1) that the scores of the neurotic simulations

[25] S R Hathaway, and J C McKinley, "A Multiphasic Personality Schedule (Minnesota), Construction of the Schedule," *Jour Psychology,* 1940, pp 249–254, and *The Minnesota Multiphasic Personality Inventory,* University of Minnesota Press, 1943, and 1951 revised, p 5

[26] S R Hathaway, and E D Monachesi, *Analyzing and Predicting Juvenile Delinquency with the MMPI,* University of Minnesota Press, pp 15–16

[27] H Gough, "On Making a Good Impression," *Jour Educational Research,* September, 1952, pp 33–42

[28] "Simulated Patterns on the MMPI," *Jour Abn & Soc Psych ,* April, 1947, pp 215–225

were, in general, similar to the authentic cases Four competent judges were able to identify correctly 55 to 73 per cent of the simulated responses, and mechanical sorting identified 82 per cent, and (2) psychotic simulations were identified 91 to 100 per cent by the judges, and 82 per cent by mechanical sorting

In the same article[29] Gough comments on the detection of malingerers in an unstructured test (Rorschach),[30] where malingerers could be identified by responses of extreme cautiousness, hesitancy, and a tendency to reject cards, and by a minimization of response in general Ossipov[31] found that every malingerer is an actor who portrays an illness as he understands it. In assuming this role the malingerer goes to extremes, apparently believing that the more eccentric his behavior, the more mentally disordered he will be thought to be Responses by students to the *Strong Vocational Interest Test* were found by Bordin[32] to be simulations of certain specified occupational types (when they were acquainted with occupational groupings) even though the students were ignorant of the mechanics of scoring. Two factors seemed involved here (1) the degree of acceptance of an occupational stereotype as self-description, and (2) the degree of knowledge of the true occupational stereotype. The importance of this finding derives from the facts that occupational groupings are (1) found to be associated with personality traits and (2) they tend to form a hierarchy of social status

Finally, it should be evident that this brief account does not exhaust the list of devices invented to correct malingering We have merely chosen some examples which have been presented as clues to the resolution of some of the difficulties encountered by the research worker in his effort to evaluate personality variations by scales which depend upon verbal responses to obtain ratings on behavior

[29] *Ibid*, p 215

[30] A non-technical discussion of the Rorschach test appears in an article by L Wekstein, "X-Raying the Personality," *Scientific Monthly*, August, 1947, pp 133–142

[31] V P Ossipov, "Malingering The Simulation Psychosis," *Bull Menninger Clinic*, 1944, no 8, pp 39–42

[32] E S Bordin, "A Theory of Vocational Interests as Dynamic Phenomena," *Educational & Psychol Measurement*, 1943, pp 49–65

Appendix A

TENTATIVE NORMS OF SOCIAL STATUS SCALE

REVISED TENTATIVE NORMS—SOCIAL STATUS SCALE SCORES (1942)

1. Previous norms (1933) suggested for interpretation of scores:

Class	Ranges in Scores	
Upper middle	125–149	
Average middle	100–124	See p 378, F Stuart
Lower middle	75–99	Chapin, *Contempo-*
Workingmen's homes	50–74	*rary American Institu-*
Relief, poverty	25–49	*tions,* Harper &
Destitute	0–24	Brothers, 1935

2 Since the depression of the thirties, several studies have shown a trend to higher scores for families receiving relief and on work relief Also, studies by George A Lundberg have shown ranges in score on the homes of normal families living in a New England village Study of results procured in different parts of the country suggests a revision of the class categories to correspond to higher social status scores as follows

3 Revised tentative norms applicable in 1942 (after Lundberg)

Class		Ranges in Scores	Examples
Upper	1 Upper part	250 and over	Professional, physician
	2 Lower part	200–249	Small town banker
Middle·	1 Upper part	150–199	Factory manager
	2. Lower part	100–149	Skilled trades
Lower:	1 Upper part	50–99	Unemployed semi-skilled
	2 Lower part	0–49	Unemployed unskilled

4 Some later studies which supply new data
George A Lundberg, "The Measurement of Socioeconomic Status," *American Sociological Review,* February, 1940, pp 29–39
G. A Lundberg and Margaret Lansing, "The Sociography of Some Community Relations," *ibid ,* June, 1937, pp 318–335
G. A. Lundberg and Mary Steele, "Social Attraction Patterns in a Village," *Sociometry,* January-April, 1938, pp 375–419.
Louis Guttman, "A Revision of Chapin's Social Status Scale," *American Sociological Review,* June, 1942, pp 362–369.
(These weights were not used in Tables 1 and 3 above)

5 Distributions of scores from representative samples of urban populations
(1) Total social status score approaches a normal or symmetrical distribution, 0 to over 220 points, slightly skewed to higher scores, modal scores 80 to 135.
(2) Score on part II, condition of the living room, a J-curve, skewed to lower scores − 16 to − 10; mode at + 8

MEASUREMENT OF CONDITION OF THE LIVING ROOM OF A DWELLING

Part II Condition of Articles in Living Room

To provide some objective rating of qualitative attributes of the living room, such as "aesthetic atmosphere" or "general impression," the following additional items may be noted The visitor should check the words that seem to describe the situation Some of the weights are of minus sign, and so operate as penalties to reduce the total score of the home

18. Cleanliness of room and furnishings

a Spotted or stained (−4) . ____

b Dusty (−2) . . ____

c Spotless and dustless (+2) . . ____

19 Orderliness of room and furnishings

a Articles strewn about in disorder (−2) . ____

b Articles in place or in usable order (+2) ____

20 Condition of repair of articles and furnishings

a Broken, scratched, frayed, ripped, or torn (−4) . . ____

b. Articles or furnishings patched up (−2) ____

c Articles or furnishings in good repair and well kept (+2) ____

21 Record your general impression of good taste

a Bizarre, clashing, inharmonious, or offensive (−4) ____

b Drab, monotonous, neutral, inoffensive (−2) ____

c Attractive in a positive way, harmonious, quiet and restful (+2) ____

Score on Part II ____ Total score,[a] Parts I and II ____

[a] The entire scale, Parts I and II, was published in F Stuart Chapin, *Contemporary American Institutions*, Harper & Brothers, 1935, and W C Brown (Dubuque), 1946, pp 373–397 Penalties for use crowding on total score of parts I and II are (1) use of living room also as a dining room, deduct 6 points, (2) used also as a kitchen, deduct 9, (3) used also as a bedroom, *or* as a dining room and kitchen combined, deduct 12, (4) used also as a bedroom, dining room, and kitchen combined, deduct 15

CHAPIN'S SOCIAL STATUS SCALE WITH THE NEW WEIGHTS COMPUTED BY LOUIS GUTTMAN[1]

PART I. MATERIAL EQUIPMENT AND CULTURAL EXPRESSION OF THE LIVING ROOM OF THE HOME

	Old Weights	New Weights
1 Floor, softwood	6	8 4
hardwood	10	13 9
2 Large rug	8	5 6
3 Windows with drapes	each 2	each 3 6
4 Fireplace with 3 or more utensils	8	34 3
5 Artificial light, electric	8	12 0
kerosene	−2	−3 0
6 Library table	8	−1 0
7 Armchairs	each 8	each 4 1
8 Piano bench	4	5 2
9 Desk personal-social	8	2.3

[1] See 4 Revised Tentative Norms—Social Status Scale Scores (1942).

	Old Weights	New Weights
10 Bookcases with books	each 8	each 3 4
11. Sewing machine	− 2	2 0
12 Couch pillows	each 2	each 7
13 Alarm clock	− 2	− 5 3
14 Periodicals	each 8	each 1 6
15 Newspapers	each 8	each 7 7
16 Telephone	8	24 4
17. Radio	8	8 0

PART II CONDITION OF ARTICLES IN LIVING ROOM

	Old Weights	New Weights
18 Cleanliness of room and furnishings		
a Spotted or stained	− 4	− 19 4
b. Dusty	− 2	− 9 7
c Spotless and dustless	2	9 7
19 Orderliness of room and furnishings		
a Articles strewn about in disorder	− 2	− 19 7
b Articles in place or in usable order	2	19.7
20 Condition of repair of articles and furnishings		
a Broken, scratched, frayed, ripped, or torn	− 4	− 16 3
b Articles or furnishings patched up	− 2	− 8 1
c. Articles or furnishings in good repair and well kept	2	8.1
21 Record your general impression of good taste		
a Bizarre, clashing, inharmonious, or offensive	− 4	− 5.4
b. Drab, monotonous, neutral, inoffensive	− 2	− 2 7
c Attractive in a positive way, harmonious, quiet and restful	2	2.7

THE SOCIAL STATUS SCALE

REVISED, 1952

VISITOR'S NAME ....

SCHEDULE NUMBER.............. DATE ....

FAMILY NAME ....

STREET ADDRESS ....

TOWN AND STATE ....

RACE ....

For original standardization on reliability and validity, which still holds, see F Stuart Chapin, *Contemporary American Institutions*, Harpers, 1935, or W C Brown (Dubuque, Iowa), 1948 Mean scores by occupational groups, as of 1950, are as follows I Professional, and II Managerial and Proprietary, 177 (upper limit 607), III Clerical, 147, IV Skilled, 117, V, Semi-skilled, 87, VI Unskilled, 57

DIRECTIONS TO VISITOR

1 The following list of items is for the guidance of the recorder Not all the features listed may be found in any one home Entries on the schedules should however, follow the order and numbering indicated Weights appear (in parentheses) after the names of the items Disregard these weights in recording Only when the list is finally checked should the individual items be multiplied by these weights and the sum of the weighted scores be computed, and then only after leaving the home All information is confidential

2 Where the family has no real living room, but uses the room at night as a bedroom, or during the day as a kitchen or as a dining room, or as both, in addition to using the room as the chief gathering place of the family, please note this fact clearly and describe for what purposes the room is used Scores corrected for these uses are as follows (1) used as dining room, deduct 6 (but if a dining alcove is structurally built in, assign *no* penalty score), (2) used as kitchen, deduct 9, (3) used as bed room *or* dining room and kitchen combined, deduct 12, (4) used as bedroom, dining room, and kitchen combined, deduct 15

3. If the item you are scoring is present, enter the figure 1 on the *dotted* part of the line If more than one item is present, write 2, 3, 4, etc, as the case may be If the item is not present, enter a zero Each division of each of the twenty items should have a number (0, 1, 2, 3, 4, etc) entered opposite it *Be sure that each item gets some number* Leave no dotted line blank A blank line means an oversight on your part

4 Disregard figures in parentheses until all schedules are finished.

5 Always enter color of each family (white or Negro, specify if any other) below town and state

6 Descriptions of items in Parts I and II

(1) Hardwood floors are usually made of narrow boards Verify your own judgment by asking the occupant

(2) A large rug is one covering the entire floor area except a border of about 12 inches (Usually a 9 x 12 foot rug)

(3) A drape is a covering over a window or windows The drape may be either at the side of the window

6 Descriptions of items in Parts I and II—*continued*

or across the top Each draped window gets one point Where the drape covers more than one window, give the question as many points as there are windows covered by the drape

(4) Only a real fireplace (one in which a coal, wood, or other fire, including gas, can be built) gets a number

(6) A library table is any table not used for serving meals (Exclude such small tables as end tables, card tables, etc) If a library table is used as a personal-social desk, number it either as a library table or as a personal desk, *but not as both*

(7) Armchairs include rocking chairs with arms

(8) Piano bench a chair or stool does not get a point.

(9) See No 6 A personal social desk is a writing desk.

(10) If no bookcase is seen, be sure to ask if there is one in another room Be sure to record any bookcase, whether in living room or in any other room, *except* professional library of doctor, lawyer, clergyman, teacher

(11) Sewing machine must be in living room in order to be marked minus two points If outside living room, mark zero

(13) Alarm clock same as No 11

(14) Number of different periodicals *regularly bought* or *subscribed* to Always get answer to this question by asking

(15) Newspapers same as No 14

(16) Telephone to be recorded whether in living room or in some other room Always ask concerning the telephone if you do not see one

(17) Radio and/or television to be recorded whether in living room or in some other room Always ask regarding these

(18) It is possible for a room to be both (a) spotted or stained and (b) dusty In this case each gets *one* check

(20) "In good repair" means in good condition and not obviously patched up, therefore *only one* of the parts (a, b, or c) gets a check.

(21) *Only one* of the parts (a, b, or c) gets a check.

Reprinted, with revisions, from *The Measurement of Social Status* by F Stuart Chapin, The University of Minnesota Press, Minneapolis, 1933

Part I Material Equipment and Cultural Expression of the Living Room of the Home

1 Floor, softwood (6) ______
   hardwood (10) ______
2 Large rug or wall-to-wall carpet (8) ______
3 Windows with shades and curtains, *or* with Venetian blinds and drapes (each window 2) ______
4 Fireplace with 3 or more utensils (8) ______
5 Artificial light, electric (8) ______
   kerosene (−2) ______
6 Library table (8) ______
7 Armchairs (8 each) ______
8 Piano bench (4) ______
9 Desk personal-social (8) ______
10 Bookcases with books (8 each) ______
11 Sewing machine (−2) ______
12 Couch pillows (2 each) ______
13 Alarm clock (−2) ______
14 Periodicals (8 each) ______
15 Newspapers (8 each) ______
16 Telephone (8) ______
17 Radio (2), television (6), both (8) ______

Score on Part I ______

Part II Condition of Articles in Living Room

To provide some objective rating of qualitative attributes of the living room, such as "esthetic atmosphere" or "general impression," the following additional items may be noted The visitor should check the words that seem to describe the situation Some of the weights are of minus sign, and so operate as penalties to reduce the total score of the home

18 Cleanliness of room and furnishings
   a Spotted or stained (−4) ______
   b Dusty (−2) ______
   c. Spotless and dustless (+2) ______

19 Orderliness of room and furnishings
   a Articles strewn about in disorder (−2) ______
   b Articles in place or in usable order (+2) ______

20 Condition of repair of articles and furnishings
   a Broken, scratched, frayed, ripped, or torn (−4) ______
   b Articles or furnishings patched up (−2) ______
   c Articles or furnishings in good repair and well kept (+2) ______

21 Record your general impression of good taste
   a Bizarre, clashing, inharmonious, or offensive (−4) ______
   b Drab, monotonous, neutral, inoffensive (−2) ______
   c Attractive in a positive way, harmonious, quiet and restful (+2) ______

Score on Part II ______ Total score,* Parts I and II ______

* With penalties deducted (See paragraph 2 on the other side)

Appendix B

TENTATIVE NORMS OF SOCIAL PARTICIPATION SCALE

SOCIAL PARTICIPATION SCALE—TENTATIVE NORMS (1944)

Occupational Groups	N	Mean Social Participation Scores	Means	Range
I. Professional and business leaders	22	1 Top business and professions	(65)	46–65
	26	2. Top civic leaders	(51)	
	17	3 Emergent leaders[a]	(46)	
	24	4 Top U-student leaders	(56)	
II. Semi-professional and managerial	18	5 Proprietary and managerial	(26)	26–32
	12	6 Labor leaders	(28)	
III Clerical workers	46	7 Clerical workers	(16)	14–16
IV Skilled workers	10	8 Skilled workers	(10)	8–15
V Semi-skilled	4	9. Semi-skilled workers		13–15
Unemployed,	44	10 Slum population	(8)	
Relief	44	11 Relief clients	(5)	
Random Sample	233	12 Random sample[b] of population (includes Nos 3 and 6)	(10)	

[a] A group of natural leaders of the community who emerged from a survey on a representative urban sample F Stuart Chapin, *Community Leadership and Opinion in Red Wing*, University of Minnesota Press, 1945, 26 pp, especially pp 6, 23

[b] *Ibid*, p 24

Social Participation Scale, 1952 Edition

By F. STUART CHAPIN
University of Minnesota

This scale measures the degree of a person's or family's participation in community groups and institutions It was first published in 1928 and its use since then has given the following partial standardization, among urban families reliability coefficients r = .89 to 95, validity coefficients, with social status r = 62 to .66, with income class r = 52, with occupational groups r = 63, with years of formal education r = 54, between husband and wife r = 76

This is a Guttman type scale with reproducibility coefficients of 92 to 97 for groups of leaders High scores of 18 and over represent titular leader achievement The five components of the scale measure different dimensions. intensity of participation by Nos. 2, 3, 4, and 5, extensity by No 1 Also, acceptance-rejection in formal groups is measured by Nos 1, 4, and 5, for which the intercorrelations are found to be of the order of r_{14} = 53 to 58; r_{15} = .36 to 40, r_{45} = 36 to 40 Social participation proper is measured by Nos 2 and 3, with intercorrelations of r_{23} = 80 to 89 Other interrelationships among the components have been found to be of the order of r_{12} = 88, r_{13} = 89, r_{24} = 60, r_{25} = 60, r_{34} = 40, r_{35} = 35, and r_{45} = 50 to .58

Mean scores for occupational groups are as follows. I Professional, and II Managerial and Proprietary, 20 (upper limit 99), III Clerical, 16, IV Skilled, 12, V Semiskilled, 8, and VI Unskilled, 4

DIRECTIONS

1 List by name the organizations with which the husband and wife are affiliated (at the *present time*) as indicated by the five types of participation *No 1 to No 5 across the top of the schedule (over)*

   It is not necessary to enter the date at which the person became a member of the organization It is important to enter L if the membership is in a purely local group, and to enter N if the membership is in a local unit of some state or national organization

2 An organization means some active and organized grouping, usually but not necessarily in the community or neighborhood of residence, such as club, lodge, business or political or professional or religious organization, labor union, etc , subgroups of a church or other institution are to be included separately *provided they are organized* as more or less independent entities

3 Record under attendance the mere fact of attendance or nonattendance without regard to the number of meetings attended (corrections for the number attended *have not* been found to influence the final score sufficiently to justify such labor)

4. Record under contributions the mere fact of financial contributions or absence of contributions, and *not the amount* (corrections for amount of contributions *have not* been found to influence the final score sufficiently to justify such labor)

5. Previous memberships, committee work, offices held, etc , should *not* be counted or recorded or used in computing the final score

6 Final score is computed by counting each membership as 1, each attended as 2, each contributed to as 3, each committee membership as 4, and each office held as 5 If both parents are living regularly in the home, add their total scores and divide the sum by two The result is the mean social participation score of the family In case only one parent lives in the home, as widow, widower, etc , the sum of that one person's *participations is the score for the family* (*unless it is desired to obtain scores on children also*)

SOCIAL PARTICIPATION SCALE

Address Case No

HUSBAND

Age Education Race or Nationality

Occupation Income

Name of Organization	1 *Member* [a]	2 *Attendance*	3 *Financial Contributions*	4 *Member* of *Committees (Not Name)*	5 *Offices Held*
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
Totals					

WIFE

Age . Education Race or Nationality

Occupation . Income .

Name of Organization	1 *Member* [a]	2 *Attendance*	3 *Financial Contributions*	4 *Member of Committees (Not Name)*	5 *Offices Held*
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
Totals					

Date . . Investigator .

[a] Enter L if purely local group, enter N if a local unit of a state or national organization

Distribution of total scores from a representative sample of an urban population, a J-curve, skewed to higher scores of 100 and over, mode at 0 to 11 points

Appendix C

COMMENT ON "EXPERIMENTAL DESIGNS IN SOCIOLOGICAL RESEARCH"[1]

In the literature of social science, two uses of the term "experiment" appear first, to stand for trial and error attempts at the resolution of some problem of human relations (this is the popular usage), and second, to describe the method of precisely controlled observation used in physical science It is to be expected that my use of the term, "experimental designs in sociological research," to mean observations of social relations under conditions of control introduced by matching, in my recent book, *Experimental Designs in Sociological Research* (New York Harper & Bros, 1947), should be misunderstood and hence challenged in reviews by mathematical statisticians Kempthorne,[2] Ackoff,[3] Keyfitz,[4] and Hagood,[5] to mention a few. In fact, at least one sociologist, Hornell Hart,[6] has made substantially the same comment and suggested that the preferred term might be, "statistical comparisons with matched control groups"

The first use of the term "experiment," as a trial and error attempt to influence social relations by social action, I specifically exclude as the meaning of the nine studies summarized in my book (pp 22–28, 29–33). My interest was to illustrate the crude beginnings of efforts to observe, under conditions of control by matching, what really happened to people when such trial and error experiments, taking the form of programs of social treatment or social reform, were used to influence them It was my purpose to show that the systematic study of social action (trial and error "experiments"), is necessary if we are to appraise objectively the results often claimed for such programs

The second meaning, I have also disclaimed (pp 4–6, 26, 29, 32–33),

[1] Reprinted from *J A S A*, September, 1949, pp 400–462
[2] O Kempthorne, *J A S A*, September, 1948, pp 489–492
[3] Russell L Ackoff, *Science*, May 14, 1948, pp 509–510
[4] Nathan Keyfitz, *A J S*, November, 1948, pp 259–260
[5] Margaret Jarman Hagood, *J A S A*, June, 1949, pp 312–313
[6] Hornell Hart, *Social Forces*, October, 1948, pp 96–98

although the term "experiment" was sometimes used as an abbreviation of the more cumbersome term, "experimental designs in sociological research," but when so used its reference was made clear in the context in which it occurred, even if the original author, whose work was summarized, used the term carelessly (Chap. 4)

These explanations and qualifications still leave us with the need of a term which may be used to describe studies of problems of social relations in which (1) a group of persons who receive a program of treatment is compared with (2) a group excluded from this treatment, the situation is, (3) such comparisons take place in the natural community (not in the artificial situation of the laboratory or in the class room situation), and finally, (4) embrace an attempt to control by matching some of the factors in the situation other than (a) the pattern of treatment factors, and (b) the pattern of response factors Studies of this sort I have called experimental designs in sociological research They are not, of course, "experiments" in Fisher's meaning of the term, and I have made no such claim

The studies which I summarized were conducted in the complex community situation—hence the qualifying terms, "in sociological research." I suspect that the full import of these latter words can hardly be meaningful to any one who has not done empirical research in this field of complex social forces, although several reviewers [7] seem quite aware of their significance I am still of the opinion that the designation, "experimental designs in sociological research" is a more meaningful descriptive title for these studies than Greenwood's [8] single term "experiment" (that promises too much) which he applied to several of these studies, or Hart's [9] "controlled comparisons" (which promises too little). No doubt usage will determine the acceptance and survival of these terms, as it does all other terms

Inadequate definitions of the universes sampled is another point of criticism As to this point, several of the investigators whose work I described did not, unfortunately, offer any such definitions The chief explanation appears to be that the purpose of each study was to observe

[7] See reviews by Hagood, *op cit*, R F Sletto, *Int J Opinion & Att Res*, Fall, 1948, pp 412–413, Donald E Super, *J Appl Psychol*, February, 1949, pp 93–95, T G Andrews, *Psychometrika*, December, 1948, pp 281–283, and Otis D Duncan, *Rural Sociology*, June, 1948, p 199

[8] Ernest Greenwood, *Experimental Sociology*, King's Crown Press, 1945, pp. xv, 163

[9] Hart, *op cit*, p 98

what happened to specific groups of subjects when treatment was or was not applied. The studies were exploratory and none was designed to provide a basis for generalization to a universe Since, however, the criticism is well taken, it may be worthwhile to note in passing that there are at least two types of universe concerned where social attitudes are measured First, there is the universe of respondents in a definite area at a given time Area sampling technique is the preferred procedure to be followed in all such cases But two hypothetical universes are also involved, and as yet little is known about the sampling techniques to be employed in such cases One of these universes is that of all possible variations in the attitude of each person on a given issue over a period of time This universe may be further subdivided into public attitudes and private attitudes, a problem of considerable practical as well as theoretical importance Then there is the universe of all possible questions that may be formulated to elicit the true attitude of a person on a given issue Stouffer[10] has recognized this problem by his statement, "Compared with the sampling of respondents, the developments in the sampling of items are still in a relatively primitive state" These points illustrate the great complexity of the concept of universe in sociological research

The impossibility of randomization of the treated group of persons and the untreated group presents the most serious problem and is a well-deserved criticism of randomization that has been neglected through ignorance; but randomization was not possible as fully explained in considerable detail (pp 168–169) To state the problem briefly, no administrator of a private treatment program or of a legalized public reform effort would permit the recipients of treatment to be chosen at random The local mores of the community or the public law that governs the program determines eligibility for treatment, whether it be by social case work, relief or public housing, by *need* for such treatment In all such studies of programs we deal with a *de facto* situation which the observer cannot control It would be an important contribution to research if some way could be found to get around this serious limitation Furthermore, the consequences of such a limitation both as it precludes generalization from samples to a universe, and as it precludes control of unknown factors, is also discussed in detail (pp 60, 78, 83, 89–90, 140–141, 166–169, 179–186) so that this point was by no means neglected It

[10] S A Stouffer, "Government and the Measurement of Opinion, *Scientific Monthly,* December, 1946, pp 435–440, especially p 436

may be also worth noting that because randomization was not possible no effort was made to use the powerful tool of analysis of variance This restriction seemed to the author such an obvious logical consequence that no explanation of detailed reasons seemed necessary, although by implication, at least, the reviews by mathematical statisticians suggest use of analysis of variance Perhaps the best statement on this point is still that made by Margaret Hagood [11] in 1941, "Are the methods of analysis of variance (and covariance) applicable for observational situations where data are observed in the cross classifications in which they are found rather than in cross classifications in which they have been randomly placed in an experiment?"

Since the sociologist seems at this stage of development of his field research to be denied the advantages of analysis of variance, resort is made of necessity to the convergence of such evidence as he has (1) the occurrence of small differences that are in the same direction of logical meaning (pp 42–46, 49, 103–106), and (2) the persistence of such differences in more and more homogeneous samples after losses from mobility, death, refusals and inability to match have taken their toll Patterns of differences which still persist after extreme cases have been lost are regarded as evidence of real differences, and hence may be taken as evidence of the effects of treatment, subject always to the qualifications set by absence of control of unknown factors

Questions naturally arise as to the validity of applying the conventional tests of statistical significance to differences between the statistics of samples that are non-random Here again, I can but refer the reader to a somewhat detailed discussion of this point (pp 176–186) wherein the qualifications and limitations are examined with the conclusion that only empirical results can be claimed, or as one reviewer puts it my solution is "a pragmatic rather than a fully analytic solution" [12]

My concept of an *ex post facto* experimental design will have to rest on evidence yet to be gathered Whether it will weather the tests of results in future research only time can tell.

Differences of opinion on the meaning of the concepts "null hypothesis" and "causation" appear in the literature of statistics and of logic As to the null hypothesis my use follows that of Lindquist,[13] who appears

[11] *Statistics for Sociologists,* Reynal & Hitchcock, Inc , 1941, pp 586 See also pp. viii, 934

[12] Duncan, *op cit*

[13] E F Lindquist, *Statistical Analysis in Educational Research,* Houghton-Mifflin Co , 1940, p 15 See also pp xiii, 266

to follow an earlier edition of Fisher's *Design of Experiments,* in which the concept was used to denote any exact hypothesis that one may be interested in disproving, and not merely the hypothesis that a certain parameter is zero, which later has come to be the conventional meaning among some groups of statisticians Some other statisticians seem also to use the meaning I attach to the term [14] For the sociologist, whose universes are complex and difficult to define, parameter values are seldom known, but there are real advantages in negative statement of relationships with the object of disproving them, because as I try to explain (pp 70–73, 83–84, 93, 137–138, 167, 187) such formulations help us get rid of normative considerations Criticisms of my use of the term seem to be based upon the contention that it departs from the current conventional usage in biological experiments and in biometrics. There are, however, examples of different usages of the same technical concept or term in other areas of research For example, the term "ambivalence" has one meaning in chemistry and a different but accepted meaning in psychiatry, and the term "rationalization" has one meaning in systematic economics and a different meaning in abnormal psychology

As to my use of the concept "causation," it will be noted in the first place that no proof of causation is claimed for any of the studies summarized (pp 50, 74, *passim*), that replication of each study is recommended as the next step toward fuller understanding of the relationships (pp 31, 57, 90, 120, 139, 176, 177, 185, 188, 189), and that my definition of causation (pp 52–54) is consistent with the position of logicians as represented in the recent book of Herbert Feigl and Wilfrid Sellers, *Readings in Philosophical Analysis* (New York Appleton-Century-Crofts, Inc, 1949)

[14] Hagood, *op cit*, p 359, and G W Snedecor, *Statistical Methods,* Collegiate Press, Inc, 1937, p 54 See also pp xiii, 341

Appendix D

TABLE A Haberman's Table of Probability Values for Random Samples [1]

S^a	Frequencies of S in 2 × 2, 2 × 3, etc , Tables [c]				
	2 × 2 [b]	2 × 3	2 × 4	2 × 5	2 × 6
0	6	90	2,520	113,400	7,484,400
1	6	90	2,520	113,400	7,484,400
2	6	108	3,240	151,320	10,206,720
3	4	114	3,840	188,880	13,153,680
4	2	108	4,272	225,000	16,254,000
5		90	4,464	256,320	19,368,720
6		60	4,392	281,040	22,389,120
7		36	4,008	296,400	25,153,200
8		18	3,336	300,960	27,537,840
9		6	2,664	292,560	29,373,120
10			1,992	270,480	30,538,080
11			1,368	244,560	30,877,200
12			864	213,720	30,289,680
13			480	179,880	29,264,400
14			240	145,920	27,679,680
15			96	113,400	25,612,560
16			24	84,480	23,195,520
17				59,880	20,533,680
18				40,560	17,774,640
19				25,920	15,019,200
20				15,480	12,398,400
21				8,520	9,977,760
22				4,200	7,840,080
23				1,800	5,995,440
24				600	4,452,480
25				120	3,204,000
26					2,228,400
27					1,492,560
28					958,320
29					587,520
30					340,560
31					184,320
32					91,440
33					40,320
34					15,120
35					4,320
36					720
TOTALS	24	720	40,320	3,628,800	479,001,600

[1] Notes for Table A, Appendix D, are on p. 285

NOTES FOR TABLE A

[a] The mean S-values for any column are obtained by the formula

$$\bar{S} = \frac{m(5 + 7m)}{24}, \text{ where } m = \text{the number of pairs}$$

[b] The probability is obtained by the sum of frequencies up to and including the table value of S, divided by the total

For example, where $S = O$, the probability is $\frac{6}{24} = \frac{1}{4}$,

or 25

where $S = 1$, the probability is $\frac{6+6}{24} = \frac{1}{2}$,

or 50

[c] For non-random samples, the Haberman Table B is not available herein, but will be found in his doctor's dissertation, *Special Distributions of Kendall's Tau* (τ), University of Minnesota, 1955, or *Biometrika*, 1956

Appendix E

LIST OF PERIODICAL ABBREVIATIONS

J.A S A.	Journal of the American Statistical Association
A J S	American Journal of Sociology
A.S R.	American Sociological Review

INDEX OF NAMES

INDEX OF SUBJECTS